

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 13:05:01 ; Search time 519.52 Seconds
(without alignments)
531.373 Million cell updates/sec

Title: US-09-867-034-2

Perfect score: 322

Sequence: 1 gcaagcaacaacacctga.....ctccaagtgtcgaagact 322

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No:	Score	Query Match Length	ID	Description
1	321	99.7	525 19	AAV29048
2	321	99.7	866 19	AAV29047
3	321	99.7	875 19	AAV19155
4	321	99.7	875 22	AAV63314
5	321	99.7	881 21	AAV59320
6	321	99.7	920 21	AAV58380
7	321	99.7	943 22	AAV35026
8	321	99.7	1018 20	AAZ77486
9	321	99.7	1020 20	AAZ52865
10	321	99.7	1689 20	AAZ40803
11	310	96.3	529 22	AAH55526

12	309	96.0	778	22	AAFA4884	Human breast cance
13	302.2	93.9	940	20	AAZ40846	Secreted protein E
14	275	85.4	386	19	AAV19188	Human XAG growth f
15	275	85.4	386	22	AAV63347	HuXAG-1 related ES
16	263.4	81.8	401	21	AAV7753	cDNA encoding huma
17	263.4	81.8	401	22	AAI28491	Colon tumour relat
18	249	77.3	409	21	AAV7903	cDNA encoding huma
19	249	77.3	409	22	AAI28641	Colon tumour relat
20	246.8	76.6	793	21	AAZ24578	Human lung tumor a
21	246.8	76.6	793	21	AAV65817	Human lung cancer-
22	233	72.4	525	19	AAV59329	Degenerate DNA seq
23	205.8	63.9	506	19	AAV19159	Human XAG growth f
24	205.8	63.9	506	22	AAV63334	HuXAG-1 related cd
25	188	58.4	468	21	AAV00115	Human secreted pro
26	167.8	52.1	321	22	AAV17730	Human breast cance
27	159.4	49.5	235	22	AAV68475	Human lung tumour
28	159.4	49.5	301	21	AAV7841	cDNA encoding huma
29	159.4	49.5	301	22	AAI28579	Colon tumour relat
30	157.2	48.8	501	22	AAV54130	Breast cancer prot
31	157.2	48.8	572	20	AAV88876	EST clone RT1404.
32	157.2	48.8	640	22	AAI29081	Colon tumour relat
33	157.2	48.8	728	19	AAV68999	Nucleotide sequenc
34	157.2	48.8	732	19	AAV19157	Human XAG growth f
35	157.2	48.8	732	22	AAV63316	Human huXAG-3/CCSG
36	157.2	48.8	749	21	AAV54129	Breast cancer prot
37	153	47.5	404	21	AAV65978	Human lung cancer-
38	151	46.9	398	20	AAV41145	Human secreted pro
39	149.4	46.4	420	22	AAV98692	Human ovarian canc
40	148	45.0	520	21	AAZ42286	Human 5' EST isolat
41	146.2	44.8	896	22	AAH33168	Human colon cancer
42	144.4	44.8	171	22	AAV1924	Human breast cance
43	136	42.2	462	21	AAV7985	cDNA encoding huma
44	136	42.2	462	22	AAI28723	Colon tumour relat
45	95.8	29.8	401	22	AAI29090	Colon tumour relat

ALIGNMENTS

RESULT 1	AAV29048	standard; cDNA; 525 BP.
ID	AAV29048	
XX	AAV29048	
AC	AAV29048	
XX	21-AUG-1998	(first entry)
XX	Open reading frame	human protein comprising secretory signal 9.
DE	Human protein; secretory signal; nutritional source; cytokine;	
XX	immunity; haematopoiesis; activin; inhibin; tumour; chemotactic;	
KW	chemokine; thrombolytic; anti-inflammatory; inhibition;	
KW	stomach cancer cell; ds.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..525
FT		/*tag= a
FT		/product= "human protein comprising secretory signal"
PN	W09811217-A2.	
XX		
PD	19-MAR-1998.	
XX		
PF	12-SEP-1997;	97WO-JP03239.
XX		
PR	13-SEP-1996;	96JP-0243060.
XX		
PA	(PROT-) PROTEGENE INC.	
XX	(SAGA) SAGAMI CHEM RES CENTRE.	

PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
 XX
 DR WPI; 1998-207380/18.
 DR P-PSDB; AAM37872.
 XX
 PT Human proteins with secretory signal sequences - used to treat
 PT immune deficiencies, infections, tumours, and haematopoietic
 PT disorders, etc.
 XX
 PS Claim 3; Pages 88; 131pp; English.
 XX
 CC This is the nucleotide sequence of the open reading frame of a novel
 CC human protein comprising a secretory signal (AAV29047), isolated from
 CC stomach cancer cells. Its proteins can be used as nutritional sources
 CC or supplements. The proteins may also have cytokine functions,
 CC immune modulating functions, haematopoiesis regulating activity,
 CC activin/inhibin regulating activity, chemotactic/chemokinetic activity,
 CC haemostatic and thrombolytic activity, receptor/ligand activity,
 CC anti-inflammatory activity, tumour inhibition activity.
 CC
 SQ Sequence 525 BP; 162 A; 131 C; 103 G; 129 T; 0 other;

Query Match 99.7%; Score 321; DB 19; Length 525;
 Best Local Similarity 99.7%; Pred. No. 2.3e-86;
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gacaagcaaaaccccttgatgattatcactcacttgatgagtgccacacagtcgaagc 60
 |||||||
 DB 198 gacaagcaaaaccccttgatgattatcactcacttgatgagtgccacacagtcgaagc 257
 |||||||
 OY 61 tttaagaaagtgcttgcgtgaataaagaatccagaatctggcagagcagttgtcct 120
 |||||||
 DB 258 tttaagaaagtgcttgcgtgaataaagaatccagaatctggcagagcagttgtcct 317
 |||||||
 OY 121 cctcaatctggttatgaaacaactgacaacaccttctccgtgagtgccagtatgtccc 180
 |||||||
 DB 318 cctcaatctggttatgaaacaactgacaacaccttctccgtgagtgccagtatgtccc 377
 |||||||
 OY 181 cagatattgttctgtagccatctcgacagttagagccgatatccactggaagatattc 240
 |||||||
 DB 378 cagatattgttctgtagccatctcgacagttagagccgatatccactggaagatattc 437
 |||||||
 OY 241 aaancgtctatgctacgacactgcagatagacgtctgttgccttacacatgaaagaa 300
 |||||||
 DB 438 aaacgctctatgctacgacactgcagatagacgtctgttgccttacacatgaaagaa 497
 |||||||
 OY 301 agctctcaagttgctgaagact 322
 |||||||
 DB 498 agctctcaagttgctgaagact 519
 |||||||

RESULT 2
 AAV29047
 ID AAV29047 standard; cDNA; 866 BP.
 XX
 AC AAV29047;
 XX
 DT 21-AUG-1998 (first entry)
 XX
 DE Human protein comprising secretory signal nucleotide sequence 9.
 XX
 KW Human protein; secretory signal; nutritional source; cytokine;
 KW immunity; haematopoiesis; activin; inhibin; tumour; chemotactic;
 KW chemokinetic; thrombolytic; anti-inflammatory; inhibition;
 KW stomach cancer cell; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 73..600
 FT /tag= a
 FT /product= "human protein comprising secretory

FT
 XX
 XX
 XX
 PD WO9811217-A2.
 PD 19-MAR-1998.
 XX
 XX
 PF 12-SEP-1997; 97WO-JP03239.
 PF 13-SEP-1996; 96JP-0243060.
 XX
 XX
 PA (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENTRE.
 XX
 PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
 XX
 DR WPI; 1998-207380/18.
 DR P-PSDB; AAM37872.
 XX
 PT Human proteins with secretory signal sequences - used to treat
 PT immune deficiencies, infections, tumours, and haematopoietic
 PT disorders, etc.
 XX
 PS Claim 4; Pages 114-116; 131pp; English.
 XX
 CC This is the nucleotide sequence of a novel human protein comprising
 CC a secretory signal isolated from stomach cancer cells. Its proteins
 CC can be used as nutritional sources or supplements. The proteins may
 CC also have cytokine functions, immune modulating functions,
 CC haematopoiesis regulating activity, activin/inhibin regulating
 CC activity, chemotactic/chemokinetic activity, haemostatic and
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, tumour inhibition activity.
 CC
 SQ Sequence 866 BP; 275 A; 189 C; 178 G; 224 T; 0 other;

Query Match 99.7%; Score 321; DB 19; Length 866;
 Best Local Similarity 99.7%; Pred. No. 2.7e-86;
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gacaagcaaaaccccttgatgattatcactcacttgatgagtgccacacagtcgaagc 60
 |||||||
 DB 270 gacaagcaaaaccccttgatgattatcactcacttgatgagtgccacacagtcgaagc 329
 |||||||
 OY 61 tttaagaaagtgcttgcgtgaataaagaatccagaatctggcagagcagttgtcct 120
 |||||||
 DB 330 tttaagaaagtgcttgcgtgaataaagaatccagaatctggcagagcagttgtcct 389
 |||||||
 OY 121 cctcaatctggttatgaaacaactgacaacaccttctccgtgagtgccagtatgtccc 180
 |||||||
 DB 390 cctcaatctggttatgaaacaactgacaacaccttctccgtgagtgccagtatgtccc 449
 |||||||
 OY 181 cagatattgttctgtagccatctcgacagttagagccgatatccactggaagatattc 240
 |||||||
 DB 450 cagatattgttctgtagccatctcgacagttagagccgatatccactggaagatattc 509
 |||||||
 OY 241 aaancgtctatgctacgacactgcagatagacgtctgttgccttacacatgaaagaa 300
 |||||||
 DB 510 aaacgctctatgctacgacactgcagatagacgtctgttgccttacacatgaaagaa 569
 |||||||
 OY 301 agctctcaagttgctgaagact 322
 |||||||
 DB 570 agctctcaagttgctgaagact 591
 |||||||

RESULT 3
 AAV19155
 ID AAV19155 standard; cDNA; 875 BP.
 XX
 AC AAV19155;
 XX
 DT 28-AUG-1998 (first entry)
 XX

DE	Human XAG growth factor huXAG-1 cDNA.
XX	
KW	huXAG-1; XAG; growth factor; colon cancer-specific gene;
KW	tumour marker; breast disease; liver disease; lung disease;
KW	emphysema; wound healing; diagnosis; therapy; human; ds.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Location/Qualifiers
FT	CDS
FT	71..598
FT	/*tag= a
FT	71..130
FT	/*tag= b
FT	mat_peptide
FT	131..595
FT	/*tag= c
XX	
PN	WO9807749-A1.
XX	
PD	26-FEB-1998.
XX	
XP	22-AUG-1997; 97WO-US14139.
XX	
PR	23-AUG-1996; 96WO-US13766.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PL	Dillon PJ, Ebner R, Endress GA, Yu G;
XX	
DR	WPI: 1998-169093/15.
DR	P-PSDB; AAM37844.
XX	
PT	New isolated human XAG growth factor(s) - used to develop products
PT	for treating e.g. liver, lung or breast diseases or
PT	hyperproliferative disorders, e.g. cancer.
XX	
PS	Claim 5; Fig 1; 141P; English.
XX	
CC	This cDNA clone codes for huXAG-1 (see AAM37844), a member of a novel
CC	family of human growth factors also including huXAG-2 (see AAM37845)
CC	and huXAG-3 (see AAM37846). These proteins share homology with the
CC	XAG protein of Xenopus laevis, which is involved in embryogenesis
CC	and is expressed in adult tissue. Expression of huXAG-1 has been
CC	discovered in colon cancer tissue, with no corresponding expression
CC	in healthy colon tissue. The huXAG-1 gene, also designated
CC	colon cancer-specific gene (CCSG), provides a molecular marker for
CC	colon cancer. huXAG-1 cDNA was isolated from a cDNA library
CC	derived from human colon cancer tissue. Vectors, host cells,
CC	antibodies, and screening methods for identifying agonists and
CC	antagonists of huXAG-1 are provided. HuXAG polypeptides are growth
CC	factors and can be used to stimulate proliferation of cells. They
CC	can be used to stimulate the proliferation and differentiation of
CC	hepatocytes to alleviate or treat liver diseases and pathologies
CC	such as fulminant liver failure caused by cirrhosis, liver damage
CC	caused by viral hepatitis and toxic substances. They can also be
CC	used to stimulate or promote liver regeneration, e.g. after
CC	surgery. They can also be used to prevent and heal damage to the
CC	lungs caused by various pathological states. They can be used to
CC	stimulate proliferation and differentiation and promote the repair
CC	of alveoli and bronchiolar epithelium to prevent, attenuate, or
CC	treat acute or chronic lung disease, e.g. emphysema, which results
CC	in the progressive loss of alveoli, and inhalation injuries, e.g.
CC	resulting from smoke inhalation and burns, that cause necrosis of
CC	the bronchiolar epithelium and alveoli. They can also be used to
CC	stimulate the proliferation and differentiation of breast tissue
CC	and could therefore be used to promote healing of breast tissue
CC	injury due to surgery, trauma or cancer. Antagonists can be used
CC	to treat hyperproliferative disorders, including cancer, in
CC	particular hepatocellular carcinoma, osteoclastoma, breast cancer,
CC	or colon cancer. The products can also be used for detection and
CC	diagnosis.
XX	
Sequence	875 BP; 283 A; 189 C; 180 G; 223 T; 0 other;

Query March	99.7%	Score 321:	DB 19:	Length 875:
Best Local Similarity	99.7%:	Pred. No.2.7e-86:		
Matches 321:	Conservative 0:	Mismatches 1:	Indels 0:	Gaps 0:
QY	1	gacaagcaacaacacctgatactatattcatcaacttgatgagtgagtgccacagctcaagc	60	
Db	268	gacaagcaacaacacctgatactatattcatcaacttgatgagtgagtgccacagctcaagc	327	
QY	61	tttaagaagaagtggttgcgtgaataataacgaagaattgscagagcagttgtccct	120	
Db	328	tttaagaagaagtggttgcgtgaataataacgaagaattgscagagcagttgtccct	387	
QY	121	cctcaatctggtgttttgtaaacacagcaacaaaccttctcctgtagtgcagatagtcctc	180	
Db	388	cctcaatctggtgttttgtaaacacagcaacaaaccttctcctgtagtgcagatagtcctc	447	
QY	181	caggaattatgtttgttgaccacactctctgaagctttagagccgatatcaactggaagaatattc	240	
Db	448	caggaattatgtttgttgaccacactctctgaagctttagagccgatatcaactggaagaatattc	507	
QY	241	aaanagctctatgcttcttacaacctgcagatacagctctgtgtgcttgcacacataagaaga	300	
Db	508	aaatgcctctatgcttcttacaacctgcagatacagctctgtgtgcttgcacacataagaaga	567	
QY	301	agctctcaagttgctgaagact 322		
Db	568	agctctcaagttgctgaagact 569		

CC	XX	RESULT 4
CC	XX	AAf63314
CC	XX	ID AAF63314 standard; CDNA; 875 BP.
CC	XX	AAf63314;
CC	XX	04-MAY-2001 (first entry)
CC	XX	Human huxAG-1/CCSG colon cancer specific gene cDNA.
CC	XX	Human; growth factor; huxAG-1; colon cancer specific gene; CCSG;
CC	XX	cell proliferation; liver disease; fulminant liver failure; cirrhosis;
CC	XX	hepatitis; cancer; colon cancer; colorectal carcinoma; ss.
CC	XX	Homo sapiens.
CC	XX	US6171816-B1.
CC	XX	09-JAN-2001.
CC	XX	22-AUG-1997; 97US-0916576.
CC	XX	23-AUG-1996; 96US-0024347.
CC	XX	(HUMA-) HUMAN GENOME SCI INC.
CC	XX	Yu G, Dillon PJ, Ebner R, Endress GA;
CC	XX	WPI: 2001-136872/14.
CC	XX	P-PSDB; AAB72203.
CC	XX	Novel human growth factor polypeptide useful for diagnosing and
CC	XX	treating colon cancer and liver diseases, to prevent and heal damage to
CC	XX	the lungs and for identifying modulators of therapeutic use -
CC	XX	Claim 45; Fig 1; 59pp; English.
CC	XX	This invention relates to a human growth factor polypeptide huxAG-1 also
CC	XX	known as a colon cancer specific gene (CCSG). HuxAG-1 stimulates cell
CC	XX	proliferation as a growth factor. The HuxAG-1 protein is useful for
CC	XX	identifying compounds capable of enhancing or inhibiting cellular
CC	XX	response induced by huxAG-1. The protein is also useful for stimulating
CC	XX	proliferation of cells e.g. colon, breast, liver and lung cells, and

CC hepatocytes. It is useful for alleviating or treating liver diseases and
CC pathologies such as fulminant liver failure caused by cirrhosis, liver
CC damage caused by viral hepatitis and toxic substances, for preventing and
CC treating damage to the lungs caused by various pathological strains and
CC for promoting healing of breast tissue injury due to surgery, trauma or
CC cancer. huxAG-1 and the identified antagonist are useful for treating
CC cancer, in particular colon cancer. Detecting altered levels of huxAG-1
CC and its polynucleotides are useful for diagnosing or detecting cancer in
CC mammals. The gene encoding huxAG-1 is useful for monitoring human
CC colorectal carcinoma. huxAG-1 nucleic acid molecules are also useful for
CC chromosome identification. The present sequence represents cDNA encoding
CC huxAG-1.
XX
XX

SQ Sequence 875 BP; 283 A; 189 C; 180 G; 223 T; 0 other;

Query Match 99.7%; Score 321; DB 22; Length 875;
Best Local Similarity 99.7%; Pred. No. 2.7e-86;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagacaacacccttgatgattatcacttgatgagtgccacacagtcacg 60
DB 268 gacaagacaacacccttgatgattatcacttgatgagtgccacacagtcacg 327
QY 61 tttaagaaagtgcttgcaaaataagaatccagaatctggcagagcagttgtcct 120
DB 328 tttaagaaagtgcttgcaaaataagaatccagaatctggcagagcagttgtcct 387
QY 121 cctcaatcgtgttatgaacacactgacaacacaccttcctccgtagggccagatgccc 180
DB 388 cctcaatcgtgttatgaacacactgacaacacaccttcctccgtagggccagatgccc 447
QY 181 caggaattatgtttgtgaccatctctgacagttagagccgatacactgaaagataatc 240
DB 448 caggaattatgtttgtgaccatctctgacagttagagccgatacactgaaagataatc 507
QY 241 aaangctctctagctttagaaccctgcagatacagctctgttcttgacaacatgaagaa 300
DB 508 aaatcgtctcttagctttagaaccctgcagatacagctctgttcttgacaacatgaagaa 567
QY 301 agctccaagttgctgaagact 322
DB 568 agctccaagttgctgaagact 589

RESULT 5
AAVS9320
ID AAVS9320 standard; cDNA; 881 BP.
XX
AC AAVS9320;
XX

DF 21-DEC-1998 (first entry)
XX

DE Nucleotide sequence encoding zsig10 polypeptide.
XX

KW SS; human; mucous-mediated function; adhesion; tumour metastasis;
KW bacterial colonisation; microbial infection; AIDS; cystic fibrosis;
KW chronic obstructive pulmonary disease; asthma; Crohn's disease;
KW sinonasal inflammatory disease; inflammatory bowel disease; bronchitis.
XX

OS Homo sapiens.
XX

XX Key Location/Qualifiers
FH CDS 63..590
FT /*tag= a
FT /product= "zsig10"
XX

XX WO9841627-A1.
XX

XX 24-SEP-1998.
XX

XX 18-MAR-1998; 98WO-US05251.
XX

PR 19-MAR-1997; 9705-0039631.
XX

XX (ZYMO) ZYMOGENETICS INC.
XX

PI Sheppard PO;
XX

DR WPI: 1998-531566/45.
XX

DR P-PSDB; AAW77365.
XX

PT New isolated mucous-associated polypeptide, zsig10 - used to develop
PT products for treating e.g. tumour metastasis, microbial infections,
PT cystic fibrosis, asthma, bronchitis or inflammatory bowel disease
XX

PS Claim 37; Page 80-81; 109pp; English.
XX

CC The human polypeptide zsig10 is involved in mucous-mediated functions
CC such as adhesion. The products of the invention can be used in the study
CC and treatment of e.g. tumour metastasis, bacterial colonisation,
CC susceptibility to and persistence of infection, microbial infections,
CC AIDS, cystic fibrosis, chronic obstructive pulmonary disease, asthma,
CC sinonasal inflammatory disease, inflammatory bowel disease, bronchitis,
CC or Crohn's disease. The products can also be used for detection,
CC diagnosis and drug screening.
XX

SQ Sequence 881 BP; 293 A; 185 C; 180 G; 223 T; 0 other;

Query Match 99.7%; Score 321; DB 19; Length 881;
Best Local Similarity 99.7%; Pred. No. 2.8e-86;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagacaacacccttgatgattatcacttgatgagtgccacacagtcacg 60
DB 260 gacaagacaacacccttgatgattatcacttgatgagtgccacacagtcacg 319
QY 61 tttaagaaagtgcttgcaaaataagaatccagaatctggcagagcagttgtcct 120
DB 320 tttaagaaagtgcttgcaaaataagaatccagaatctggcagagcagttgtcct 379
QY 121 cctcaatcgtgttatgaacacactgacaacacaccttcctccgtagggccagatgccc 180
DB 380 cctcaatcgtgttatgaacacactgacaacacaccttcctccgtagggccagatgccc 439
QY 181 caggaattatgtttgtgaccatctctgacagttagagccgatacactgaaagataatc 240
DB 440 caggaattatgtttgtgaccatctctgacagttagagccgatacactgaaagataatc 499
QY 241 aaangctctctagctttagaaccctgcagatacagctctgttcttgacaacatgaagaa 300
DB 500 aaatcgtctcttagctttagaaccctgcagatacagctctgttcttgacaacatgaagaa 559
QY 301 agctccaagttgctgaagact 322
DB 560 agctccaagttgctgaagact 581

RESULT 6
AAC58380
ID AAC58380 standard; cDNA; 920 BP.
XX

AC AAC58380;
XX

DF 29-JAN-2001 (first entry)
XX

DE Human PRO1030 nucleotide sequence SEQ ID NO:39.
XX

KW Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW proliferation; tumorigenesis; identification; cancer; cytostatic;
KW neutropenic; neuroprotective; antiinflammatory; immunosuppressive;
KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
KW hypothalamic disorder; glandular disorder; macrophagal disorder;
KW epithelial disorder; stromal disorder; blastocoeleic disorder;
KW

KW Inflammatory disorder; immunologic disorder; ss.
XX Homo sapiens.
OS
XX WO200053755-A2..
PN
XX
PD 14-SEP-2000.
XX
XX 06-JAN-2000; 2000WO-US00376.
PF
XX
PR 08-MAR-1999; 99WO-US05028.
PR 02-JUN-1999; 99WO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 26-JUL-1999; 99US-0145698.
PR 30-NOV-1999; 99WO-US28313.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
XX
XX (GETH) GENENTECH INC.
PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;
PI Matanabe CK, Wood WT;
DR WPI: 2000-572270/53.
DR P-PSDB; AAB24070.
XX
XX Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT treatment, diagnosis and prevention of cancer -
XX
XX Claim 50; Fig 27; 286pp; English.
XX
XX The present invention describes an isolated antibody that binds to
CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO335,
CC PRO619, PRO717, PRO805, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC PRO1187, PRO1281, PRO1283, PRO1309, PRO1317, PRO1317, PRO2094,
CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC growth. The PRO polypeptides and nucleotides are useful in the
CC treatment, diagnosis and prevention of cancer. The antibodies and other
CC anti-tumour compounds may be used to treat various conditions, including
CC those characterised by overexpression and/or activation of the amplified
CC PRO genes. Exemplary conditions or disorders to be treated with such
CC antibodies and other compounds include benign or malignant tumours
CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC leukemias and lymphoid malignancies, other disorders such as neuronal,
CC glioma, astrocytoma, hypothalamic and other glandular, macrophagal,
CC epithelial, stromal and blastocoeleic disorders, and inflammatory,
CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC primers and hybridisation probes used in the isolation of the human PRO
CC sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
XX the present invention.
XX
XX Sequence 920 BP; 296 A; 198 C; 195 G; 231 T; 0 other;
SQ

Query Match 99.7%; Score 321; DB 21; Length 920;
Best Local Similarity 99.7%; Pred. No. 2.8e-86;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX

QY 1 gaaagaacaaacccctgatgatcattcattcattgatgatgagcgcacacacgaacgaac 60
DB 345 gacacgaacaaacccctgatgatcattcattcattgatgatgagcgcacacacgaacgaac 404
QY 61 tttaagaagaagtgttctgctaagaataacacagaatgtgcacagcagttgtcct 120
DB 405 tttaagaagaagtgttctgctaagaataacacagaatgtgcacagcagttgtcct 464
QY 121 cctcaatctgtgttatatgaacacatgacaaacaccttctcctgatgcccagatgtctcc 180
|||||

DB 465 cctcaatctgtgttatatgaacacatgacaaacaccttctcctgatgcccagatgtcc 524
QY 181 cagatattgttctgtgacccatctctgacagttgagacgatatcactgtgaagatattc 240
DB 525 cagatattgttctgtgacccatctctgacagttgagacgatatcactgtgaagatattc 584
QY 241 aaancgtctcatgcttcaaacctgcagatcacacagctctgttgccttgacacataagaa 300
DB 585 aaatcgtctcatgcttcaaacctgcagatcacacagctctgttgccttgacacataagaa 644
QY 301 agctctcaagtgtcgtgaagact 322
DB 645 agctctcaagtgtcgtgaagact 666
|||||

RESULT 7
AAH35026 standard; cDNA; 943 BP.
XX
XX AAH35026;
AC
XX 03-SEP-2001 (first entry)
DT
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:2108.
DE
XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
XX Homo sapiens.
OS
XX WO200122920-A2.
XX
XX 05-APR-2001.
PD
XX 28-SEP-2000; 2000WO-US26524.
PF
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI: 2001-235357/24.
PT P-PSDB: AAG75621.
DR
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1; Page 3593; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 943 BP; 311 A; 199 C; 199 G; 229 T; 5 other;
SQ

CC ovarian cancer; (11) directly for treating this form of cancer

DR WPI; 1999-621386/54.
D-ECCD: 1AV73036 1AV73037

DR P-PSDB; AAY73835, AAY73836, AAY73837.

XX	New human nucleic acid sequences from pancreatic tumors, and related proteins -
PT	
PS	Claim 2; Page 188; 502pp; German.
XX	
CC	This invention describes novel polypeptides and their encoding nucleic acids derived from human pancreatic tumor tissue which have cytostatic activity. The sequences are also useful in producing pharmaceutical compositions for treatment of pancreatic tumors. AAZ52858-253014
CC	represent expressed sequence tag (EST) fragments derived from a human pancreatic tumor cDNA library and which encode the proteins represented in AAT73814-Y74252.
CC	
XX	
SQ	Sequence 1020 BP; 341 A; 214 C; 218 G; 247 T; 0 other;
Query Match	99.7%; Score 321; DB 20; Length 1020;
Best Local Similarity	99.7%; Pred. No. 2.9e-86;
Matches 321; Conservative	0; Mismatches 1; Indels 0; Gaps
DG	1 gacagcaacaacacctgatgattatcattcatcttgtagtggccacacagtcaagc 60
DG	347 gacagcaacaacacctgatgattatcattcatcttgtagtggccacacagtcaagc 406
DG	61 tttaagaagaagttgttcgtcgaaataaagaataaccagaaatgtgcagagcagttgtcct 120
DG	407 tttaaagaagaagttgttcgtcgaaataaagaataaccagaaatgtgcagagcagttgtcct 466
DG	121 cctcaatctgttttatgaacaactgcagaacacctctctcctgtagtggcagtagtccc 180
DG	467 cctcaatctgttttatgaacaactgcagaacacctctctcctgtagtggcagtagtccc 526
DG	181 caggattatgtttgttgaccatctctbcagattagagccgatalaacacgagaatatlc 240
DG	527 cagagattatgtttgttgaccatctctbcagattagagccgatalaacacgagaatatlc 586
DG	241 aaanagttctatgcttcgaacacctgcagatcacagctctgttgccttgacacatgaaga 300
DG	587 aaacgctctcatgcttacgaaacctgcagatcacagctctgttgccttgacacatgaaga 646
DG	301 agctctcaagttgctgaagact 322
DG	647 agctctcaagttgctgaagact 668
RESULT 10	
AAZ40803	
ID	AAZ40803 standard; DNA; 1689 BP.
XX	
AC	AAZ40803;
XX	
DT	18-JAN-2000 (first entry)
DE	Secreted protein EST coding sequence 108-008-5-0-A6-FL.
XX	
KW	Secreted protein; fingerprint identification technique; Chromosome mapping; human; hereditary disease; diagnosis; cancer; hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy; autoimmune disease; rheumatic disease; embryogenic disorder; myopathy; renal injury; amino aciduria; hypoglycaemia; male rat infertility; hypertension; ss.
OS	Homo sapiens.
PN	MO9940189-A2.
PD	12-AUG-1999.
PF	09-FEB-1999; 99MO-IB00282.
RR	09-FEB-1998; 98US-0074121. 13-APR-1998; 98US-0081563.

PR	10-AUG-1998:	98US-0096116.
PR	04-SEP-1998:	98US-00929273.
XX		
PA	(GEST) GENSET.	
PI	Bougueleret L, Duclert A, Dumas Mline Edwards J;	
XX		
DR	WPI; 1999-600966/51.	
DR	P-PSDB; AAY59675.	
XX		
PT	Extended cDNAs useful for expressing secreted proteins and to obtain	
PT	specific antibodies -	
PS	Claim 1; Page 168-169; 244pp; English.	
XX		
CC	This sequence encodes a human secreted protein of the invention. The	
CC	extended cDNAs (or genomic DNAs obtainable from them) may be used to	
CC	prepare PCR primers and probes. These are useful for forensic matching or	
CC	positive identification by DNA sequencing. They may also be used in	
CC	alternative fingerprint identification techniques. Antibodies against the	
CC	proteins encoded by the extended cDNAs are useful in identification of	
CC	tissue types or cell species, as well as identifying tissue specific	
CC	soluble proteins. The sequences can be used for chromosome mapping and	
CC	identification of genes associated with hereditary diseases or drug	
CC	response. Signal sequences from the cDNAs can be used in construction of	
CC	secretion vectors. Other sequences derived from the extended cDNAs can be	
CC	used to clone upstream genomic DNA sequences including promoters. This is	
CC	in turn useful for identifying proteins that interact with promoter	
CC	sequences. Some of the proteins may be useful in diagnosing and treating	
CC	several disorders including, but not limited to: cancer, hyperlipidaemia,	
CC	cardiovascular and neurodegenerative disorders, autoimmune diseases, and	
CC	rheumatic diseases, embryogenic disorders, hypertension, renal injury,	
CC	amino acidurias, hypoglycaemia, male rat infertility and myopathies.	
XX		
SO	Sequence 1689 BP; 552 A; 350 C; 335 G; 452 T; 0 other:	
Query Match	99.7%; Score 321; DB 20; Length 1689;	
Best Local Similarity	99.7%; Pred. No. 3.5e-86;	
Matches 321; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 gacaagcaacaacccctgatgatattatcattcaacttgatgagtcgccacacagtcacg 60	
DB	248 gacacagcaacaacccctgatgatattatcattcaacttgatgagtcgccacacagtcacg 307	
QY	61 tttaagaagaatggtttgttgtaaaataaagaagaatccagaaatggcagagcagttgtcct 120	
DB	308 tttaagaagaatggtttgttgtaaaataaagaagaatccagaaatggcagagcagttgtcct 367	
QY	121 cctcaatctggttttaagaacaacatcgacaacaacacctctcctgtatgagcagatgtccc 180	
DB	368 cctcaatctggttttaagaacaacatcgacaacaacacctctcctgtatgagcagatgtccc 427	
QY	181 caggaattatggtttgtgacccaactctctyagcagltagagccgatalcactcygaagatlc 240	
DB	428 caggaattatggtttgtgacccaactctctyagcagltagagccgatalcactcygaagatlc 487	
QY	241 aaanagctctatgcttgcgaacctgcgcagatatacagctctgttgccttgacaacatgaaga 300	
DB	488 aaatgcgtctcatatgcttgcgaacctgcgcagatatacagctctgttgccttgacaacatgaaga 547	
QY	301 agctctcaagttgcgtgaagact 322	
DB	548 agctctcaagttgcgtgaagact 569	
RESULT 11		
ID	AAH5526/c	
XX	AAH5526 standard; DNA: 529 BP.	
AC	AAH5526;	
XX		
DT	04-SEP-2001 (first entry)	

```
XX XX Human breast tumour protein contig 10 DNA sequence.
DE XX
XX XX Cytostatic; vaccine; human; breast tumour protein; breast cancer;
KM gene therapy; ds.
XX XX
OS Homo sapiens.
XX XX
PN WO200140269-A2.
XX XX
PD 07-JUN-2001.
XX XX
PF 29-NOV-2000; 2000MO-US32520.
XX XX
XX 30-NOV-1999; 99US-0451651.
PR 22-FEB-2000; 2000US-0510662.
PR 10-MAR-2000; 2000US-0523586.
PR 07-APR-2000; 2000US-0545068.
PR 15-MAY-2000; 2000US-0571025.
XX XX
XX (CORI-) CORIXA CORP.
XX XX
PI Dillon DC, Day CH, Jiang Y, Houghton RL, Mitcham JL, Wang A;
XX WPI; 2001-356154/37.
XX XX
PT Breast tumor polypeptides and the nucleic acids that encode them,
XX useful for the prevention, diagnosis and treatment of breast cancer -
XX
PS Claim 24; Page 138-139; 221pp; English.
XX XX
CC The present sequence is a human breast tumour protein coding sequence.
CC This sequence may be used in the prevention, diagnosis and treatment of
CC diseases associated with inappropriate expression of the breast tumour
CC protein e.g. breast cancer. For example, this sequence may be used to
CC treat disorders associated with decreased expression by rectifying
CC mutations or deletions in a patient's genome that affect the activity of
CC breast tumour protein by expressing inactive proteins or to supplement
CC the patients own production of the breast tumour protein. Additionally,
CC the present sequence may be used to produce the breast tumour protein, by
CC inserting the nucleic acids into a host cell and culturing the cell to
CC express the protein. The present sequence and its complementary sequences
CC may also be used as DNA probes in diagnostic assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy.
XX XX
SQ Sequence 529 BP; 147 A; 104 C; 102 G; 176 T; 0 other;
```

```
Query Match 96.3%; Score 310; DB 22; Length 529;
Best Local Similarity 99.4%; Pred. No. 4.4e-83;
Matches 321; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
```

```
QY 1 gacaagcaacaacccttgatga-ttatcattcattgagtgagtgccacagtcgaag 59
DB 507 GACAAGCAACAACCCCTTGATGATTTATTCATCTTGATGATGATGCCACACAGTCAAG 448
QY 60 ctttaagaagaagtggttcggaataaagaataccaagaatggcagaacagttgtcc 119
DB 447 CTTTAAAGAAAGTGTTCGCGAAATTAAGAAATCCAGAAATTTGGCAGACAGTGTCTCC 388
QY 120 tctcaatcgtgttattgaacaactgacaacaccttctcctgattgagcagatgtcc 179
DB 387 TCCTCAATCTGTTTATGAAACACTGACAAACACCTTCTCTGATGCGCATGATGTCC 328
QY 180 caaggaattatgttcttgacccatctctcgaacttagagccgatatcaactggaagatatc 239
DB 327 CCAGGATTTATGTTTGTGACCCATCTCTGACAGTTAGAGCCGATATCAGCGAAGATATT 268
QY 240 caaanctctctatgcttgaacacgcagatacagctcgtgtgcttgacaataaga 299
DB 267 CAAACGCTCTTATGCTTACGAACCTGCAGATACAGCTCTGTGCTTGGACACATGAGA 208
```

```
QY 300 aagctcgaagttgctgaagact 322
DB 207 AAGCTCTCAAGTTGCTGGAAGACT 185
```

```
RESULT 12
AAFA4884
ID AAF44884 standard; cDNA; 778 BP.
XX
AC AAF44884;
XX
DT 28-MAR-2001 (first entry)
XX
DE Human breast cancer related protein coding sequence SEQ ID NO: 40.
XX
KM Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200078960-A2.
XX
PD 28-DEC-2000.
XX
PF 23-JUN-2000; 2000MO-US17536.
XX
PR 23-JUN-1999; 99US-0140903.
PR 12-OCT-1999; 99US-0158980.
XX
PA (CORI-) CORIXA CORP.
XX
PI Yuqin J, Mitcham JL;
XX
DR WPI; 2001-041426/05.
XX
PT New polynucleotides encoding breast tumor specific proteins, useful for
XX prevention, treatment and diagnosis of breast cancer -
XX
PS Claim 25; Page 132; 165pp; English.
XX
CC The present invention provides the coding sequences for a number of
CC breast cancer related proteins. These can be used in vaccinations
CC against, diagnosis of and treatment of cancer, particularly breast
CC cancer.
XX
SQ Sequence 778 BP; 245 A; 175 C; 158 G; 193 T; 7 other;
```

```
Query Match 96.0%; Score 309; DB 22; Length 778;
Best Local Similarity 99.4%; Pred. No. 1e-82;
Matches 320; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
```

```
QY 1 gacaagcaacaacccttgatgaattatcattcattgagtgagtgccacagtcgaag 60
DB 256 gacaagcaacaacccttgatgaattatcattcattgagtgagtgccacagtcgaag 315
QY 61 tttaagaagaagtggttcggaataaagaataccaagaatggcagaacagttgtcct 120
DB 316 tttaagaagaagtggttcggaataaagaataccaagaatggcagaacagttgtcct 375
QY 121 cctcaatcgtgttattgaacaactgacaacaccttctcctgattgagcagatgtccc 180
DB 376 cctcaatcgtgttattgaacaactgacaacaccttctcctgattgagcagatgtccc 435
QY 181 caaggaattatgttcttgacccatctcgaacagttagagccgatatcaactggaagatatc 240
DB 436 caaggaattatgttcttgacccatctcgaacagttagagccgatatcaactggaagatatc 495
QY 241 aaanctctctatgcttgaacacgcagatacagctcgtgtgcttgacacataaga 300
DB 496 aaatgctctctatgcttgaacacgcagatacagctcgtgtgcttgacacataaga 554
QY 301 agctcgaagttgctgaagact 322
```

Db 555 agctctcaagttgctgaagact 576

RESULT 13

AAZ40846

ID AAZ40846 standard; DNA: 940 BP.

XX

AC AAZ40846;

XX

DT 18-JAN-2000 (first entry)

XX

DE Secreted protein EST coding sequence 78-21-1-B7-FL1.

XX

KW Secreted protein; fingerprint identification technique;

KM chromosome mapping; human; hereditary disease; diagnosis; cancer;

KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;

KM autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;

KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;

KM hypertension; ss.

XX

OS Homo sapiens.

XX

PN WO940189-A2.

XX

PD 12-AUG-1999.

XX

PF 09-FEB-1999; 99WO-IB00282.

XX

PR 09-FEB-1998; 9805-0074121.

PR 13-APR-1998; 9805-0081563.

PR 10-AUG-1998; 9805-0096116.

PR 04-SEP-1998; 9805-0099273.

XX

PA (GEST) GENSET.

XX

PI Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX

DR P-PSDB; AAY59718.

XX

PT Extended CDNAS useful for expressing secreted proteins and to obtain

PT specific antibodies -

XX

PS Claim 1; Page 226; 244pp; English.

XX

CC This sequence encodes a human secreted protein of the invention. The

CC extended CDNAS (or genomic DNAs obtainable from them) may be used to

CC prepare PCR primers and probes. These are useful for forensic matching or

CC positive identification by DNA sequencing. They may also be used in

CC alternative fingerprint identification techniques. Antibodies against the

CC proteins encoded by the extended CDNAS are useful in identification of

CC tissue types or cell species, as well as identifying tissue specific

CC soluble proteins. The sequences can be used for chromosome mapping and

CC identification of genes associated with hereditary diseases or drug

CC response. Signal sequences from the CDNAS can be used in construction of

CC secretion vectors. Other sequences derived from the extended CDNAS can be

CC used to clone upstream genomic DNA sequences including promoters. This is

CC in turn useful for identifying proteins that interact with promoter

CC sequences. Some of the proteins may be useful in diagnosing and treating

CC several disorders including, but not limited to: cancer, hyperlipidaemia,

CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and

CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,

CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.

XX

SO Sequence 940 BP; 304 A; 202 C; 187 G; 233 T; 14 other;

QY

Query Match 93.9%; Score 302.2; DB 20; Length 940;

Best Local Similarity 96.0%; Pred. No. 1.2e-80;

Matches 310; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

1 gacaagcaacaacccctgatgatcattcatcacttgatgatgagtcacacacagtcgaac 60

|||||

Db 374 racaagcaacaacacctgatgatattcatcacttgatgagtcacacacagtcgaac 433

QY 61 tttaagaagaagtgtttgtctgaaataaagaataatccagaatctggagagcagttgttc 120

Db 434 tttaaaaaaaktgtctgcgaataaaraaatccagaatctggacancagttgttcy 493

QY 121 cctcaatctgtttatgaaacaactgacaacacctctctctatgagccagatgt-cc 179

Db 494 cctcaatctgtttatgaaacaactgacaacacctctctctatgagccagatgt-cc 553

QY 180 ccagatattatgtttgtgacccatctctgacagttagaagccgatatcaactgagaatt 239

Db 554 ccmgattatgtttgttaccatctctgacagttagaagccgatatcaactgagaatt 613

QY 240 caanagctctatgacttgaacacctgacagatgacagctctgtgtgttgaacacaga 299

Db 614 caaagctctatgacttgaacacctgacagatgacagctctgtgtgttgaacacaga 673

QY 300 aagctctcaagttgctgaagact 322

Db 674 aagctctcaagttgctgaagact 696

RESULT 14

AAV19188

ID AAV19188 standard; CDNA: 386 BP.

XX

AC AAV19188;

XX

DT 28-AUG-1998 (first entry)

XX

DE Human XAG growth factor huxAG-1 related EST.

XX

KW huxAG-1; XAG; growth factor; colon cancer; tumour marker;

KW breast disease; liver disease; lung disease; emphysema;

KW wound healing; diagnosis; therapy; human; expressed sequence tag;

KW EST; ss.

XX

OS Homo sapiens.

XX

PN WO9807749-A1.

XX

PD 26-FEB-1998.

XX

PF 22-AUG-1997; 97WO-US14139.

XX

PR 23-AUG-1996; 96WO-US13766.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

PI Dillon PJ, Ebner R, Endress GA, Yu G;

XX

DR WPI: 1998-169093/15.

XX

PT New isolated human XAG growth factor(s) - used to develop products

PT for treating e.g. liver, lung or breast diseases or

PT hyperproliferative disorders, e.g. cancer.

XX

PS Claim 5; Page 111; 141pp; English.

XX

CC This expressed sequence tag (EST), deposited as GenBank Accession

CC No. AA244356, is related to a portion of huxAG-1 cDNA (see V19155),

CC which codes for a novel human growth factor (see W37844) that

CC shares homology with the XAG protein of *Xenopus laevis*. Expression

CC of huxAG-1 has been discovered in colon cancer tissue, with no

CC corresponding expression in healthy colon tissue. The huxAG-1 gene

CC provides a molecular marker for colon cancer. Vectors, host cells,

CC antibodies, and screening methods for identifying agonists and

CC antagonists of huxAG-1 are provided. huxAG polypeptides can be

CC used to stimulate proliferation and differentiation of hepatocytes

CC to alleviate or treat liver diseases and pathologies. They can

CC also be used to stimulate or promote liver regeneration, and also

CC to prevent and heal damage to the lungs caused by various

CC pathological states. They can also be used to treat acute or
CC chronic lung damage, and to stimulate the proliferation and
CC differentiation of breast tissue.

XX

SQ Sequence 386 BP; 122 A; 95 C; 74 G; 95 T; 0 other;

Query Match	85.4%	Score 275;	DB 19;	Length 386;
Best Local Similarity	99.6%	Pred. No. 1.1e-72;		
Matches 275;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

OY	1	gacaaagcaaacacccttgatattcatcctaactggatgagttgccacaagaagtcacagc	60
Db	111	gacaaagcaaacacccttgatattcatcctaactggatgagttgccacaagaagtcacagc	170
OY	61	ttaaagaagaagtgttgcctgcgaataaagaatccagaaatctgcagagcagtttgtcct	120
Db	171	tttaaaagaagaagtgttgcctgaataaagaatccagaaatctgcagagcagtttgtcct	230
OY	121	cctcaatctcgttatgaacaacaatgcgaacaccttcccgcgatggccaagtatgtccc	180
Db	231	cctcaatctcgttatgaacaacaatgcgaacaccttcccgcgatggccaagtatgtccc	290
OY	181	caggattatcttctgtacaccatcctcgaacagttlaagccgatatactactggaagaatatic	240
Db	291	caggattatcttctgtacaccatcctcgaacagttlaagccgatatactactggaagaatatic	350
OY	241	aaanagctctctatgcttaagcaaccctgcagatatcagc	276
Db	351	aaaatcgtctctatgcttaagcaaccctgcagatatcagc	386

RESULT 15
AA63347
ID AA63347 standard; cDNA; 386 BP.

DT 04-MAY-2001 (first entry)

DE HUXAG-1 related EST sequence SEQ ID 35.

KM Human, growth factor: huxA5-1; colon cancer specific gene, CC5G;
KM cell proliferation; liver disease; fulminant liver failure; cirrhosis;
KM hepatitis; cancer; colon cancer; colorectal carcinoma; EST;
KM expressed sequence tag; ss.

OS Homo sapiens.

PN US6171816-B1.

PD 09-JAN-2001

PF 22-AUG-1997; 97US-0916576.

PR 23-AUG-1996; 96US-0024347.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Yu G, Dillon PJ, Ebner R, Endress GA;

DR WPI; 2001-136872/14.

PT Novel human growth factor polypeptide useful for diagnosing and treating colon cancer and liver diseases, to prevent and heal damage to the lungs and for identifying modulators of therapeutic use -

PS Disclosure; Column 79-80; 59pp; English.

This invention relates to a human growth factor polypeptide huxag-1 also known as a colon cancer specific gene (CCSG). huxag-1 stimulates cell proliferation as a growth factor. The huxag-1 protein is useful for identifying compounds capable of enhancing or inhibiting cellular

CC responses induced by huxAG-1. The protein is also useful for stimulating
CC proliferation of cells e.g. colon, breast, liver and lung cells, and
CC hepatocytes. It is useful for alleviating or treating liver diseases and
CC pathologies such as fulminant liver failure caused by cirrhosis, liver
CC damage caused by viral hepatitis and toxic substances, for preventing and
CC treating damage to the lungs caused by various pathological states and
CC for promoting healing of breast tissue injury due to surgery, trauma or
CC cancer. huxAG-1 and the identified antagonist are useful for treating
CC cancer, in particular colon cancer. Detecting altered levels of huxAG-1
CC and its polynucleotides are useful for diagnosing or detecting cancer in
CC mammals. The gene encoding huxAG-1 is useful for monitoring human
CC colorectal carcinoma. huxAG-1 nucleic acid molecules are also useful for
CC chromosome identification. The present sequence represents an EST
CC (expressed sequence tag) related to the huxAG-1 sequence.
CC NOTE: Sequences AAF63347 - AAF63351 are stated as specifically not
CC claimed in claim 1.

SQ Sequence 386 BP; 122 A; 95 C; 74 G; 95 T; 0 other;

Query Match	85.4%	Score 275;	DB 22;	Length 386;
Best Local Similarity	99.6%	Pred. No. 1.1e-72;		
Matches 275;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0

Qy 1 gacaagcaacaaacccttgatgatattatcatcacttgatgatgtgccacacagtcaagc 60
|||||
Db 111 gacaagcaacaaacccttgatgatattatcatcacttgatgatgtgccacacagtcaagc 170

QY 61 ttctaaagaagtggttctcgtaaaaataagaataccagaattggcagagcagttgtcct 120
|||||
171 ttctaaagaagtggttctcgtaaaaataagaataccagaattggcagagcagttgtcct 230
Db

Qy 121 ccccaatctggtttatgaacaactgacaaaccttctcgtatgcccagtatgtccc 180
|||||
Db 231 ccccaatctggtttatgaacaactgacaaaccttctcgtatgcccagtatgtccc 290

QY 181 cagattatgtttgtgaccatctcgacagttagagccgatatcactcgaagatctc 240
|||||
Db 291 cagattatgtttgtgaccatctctcgacagttagagccgatatcactcgaagatctc 350

Oy 241 aaancgctcctatgtttagcaacctgcagatacagc 276
||| |||||||||||||||||||||||||
Db 351 aaatcgctctatgtcttaagaacctgcagatacagc 386

Search completed: March 30, 2002, 13:05:03
Job time: 21300 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 12:59:16 ; Search time 8456.85 seconds

(without alignments)
628.141 Million cell updates/sec

Title: US-09-867-034-2

Perfect score: 322

Sequence: 1 gcaacgcaacaaccccttga.....ctctcaagtgtcgtgaagact 322

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

GenBank:
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_om:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_in:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321	99.7	838	9 AF007791	AF007791 Homo sapi
2	321	99.7	875	6 AR123828	AR123828 Sequence
3	321	99.7	956	6 AF088867	AF088867 Homo sapi
4	321	99.7	1020	6 AX011612	AX011612 Sequence
5	321	99.7	1020	6 AX015056	AX015056 Sequence
6	321	99.7	1077	6 AF038451	AF038451 Homo sapi
7	310	96.3	529	6 AX156194	AX156194 Sequence
8	309	96.0	778	6 AX067336	AX067336 Sequence
9	275	85.4	386	6 AR123857	AR123857 Sequence
10	263.4	81.5	401	6 AX192465	AX192465 Sequence
11	262.4	81.5	758	10 AB016592	AB016592 Sequence
12	262.4	81.5	781	10 AF044262	AF044262 Mus muscu
13	249	77.3	409	6 AX192616	AX192616 Sequence
14	246.8	76.6	793	6 AR144213	AR144213 Sequence
15	205.8	63.9	506	6 AR123832	AR123832 Sequence
16	159.4	49.5	235	6 AX062783	AX062783 Sequence
17	159.4	49.5	301	6 AX192553	AX192553 Sequence
18	157.2	48.8	640	6 AX193063	AX193063 Sequence
19	157.2	48.8	732	6 AR123830	AR123830 Sequence
20	149.4	46.4	420	6 AX098187	AX098187 Sequence
21	136	42.2	462	6 AX192705	AX192705 Sequence
22	133.8	41.6	1004	5 AF025474	AF025474 Xenopus lae
23	132.8	41.2	1946	5 XU082110	XU082110 Xenopus lae
24	125.8	39.1	2057	5 XU076752	XU076752 Xenopus lae
25	95.8	29.8	401	6 AX193073	AX193073 Sequence
26	95.8	29.8	410	6 AX193110	AX193110 Sequence
27	93.4	29.0	489	6 AR123831	AR123831 Sequence
28	90.2	28.0	160274	2 AC073411	AC073411 Homo sapi
29	90.2	28.0	165616	9 AC073333	AC073333 Homo sapi
30	90.2	28.0	169554	2 AC021379	AC021379 Homo sapi
31	76.6	23.8	169554	2 AC021379	AC021379 Homo sapi
32	59	18.3	329	6 AR123833	AR123833 Sequence
33	59	18.3	329	6 AR123834	AR123834 Sequence
34	59	18.3	329	6 AR123835	AR123835 Sequence
35	59	18.3	329	6 AR123836	AR123836 Sequence
36	57.8	18.0	330	6 AR123837	AR123837 Sequence
37	53.6	16.6	1315	10 BC006857	BC006857 Mus muscu
38	46	14.3	640	6 AX136502	AX136502 Sequence
39	45.8	14.2	1391	6 AX136177	AX136177 Sequence
40	45.8	14.2	1409	9 BC008953	BC008953 Homo sapi
41	45.8	14.2	1422	9 BC001493	BC001493 Homo sapi
42	45.8	14.2	1423	6 AR123829	AR123829 Sequence
43	45.8	14.2	1452	9 AR131758	AR131758 Homo sapi
44	44.2	13.7	1450	9 BC008913	BC008913 Homo sapi
45	43	13.4	311	6 AR123862	AR123862 Sequence

ALIGNMENTS

RESULT	1	LOCUS	AF007791	DEFINITION	838 bp mRNA	PRI	14-NOV-1998
ACCESSION	AF007791	VERSION	AF007791	KEYWORDS	Homo sapiens, secreted cement gland protein XAG-2 homolog (hag-2/C)		
SOURCE	AF007791.1	ORGANISM	GI:3779196				
REFERENCE	1	AUTHORS	Thompson, D.A. and Weigel, R.J.				
TITLE	hag-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines						
JOURNAL	Biochem. Biophys. Res. Commun.	MEDLINE	99009231				
REFERENCE	2		(bases 1 to 838)				

AUTHORS Thompson, D.A.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1997) Surgery, Stanford, 1201 Welch Road, MSLS Building, Room P228, Stanford, CA 94305, USA
FEATURES Location/Qualifiers

source
1.838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="breast"
/cell_line="adenocarcinoma"
/cell_line="MCF7; estrogen receptor-expressing"
1.838
/gene="hag-2/c"
11.538
/gene="hag-2/c"
/note="similar to Homo sapiens protein hag-2/R deposited in GenBank Accession Number AF038451, and to Xenopus laevis cement gland proteins np77 deposited in GenBank Accession Number U82110 and XAG-2 deposited in GenBank Accession Number AF025474"
/codon_start=1
/product="secreted cement gland protein XAG-2 homolog"
/protein_id="AAC7358.1"
/db_xref="GI:3779197"
/translation="MEKIPVSAFLILVALSYTLARDFTYKPGAKKDTKSRPKLPOTL SRGWGQILWQTQYEEALYKSKTSNKPIMIIHLDECPSQALKKFAENKEIOKLAE OFVLNLVYETTDKHLSPDGQYVPRIMFVDPSTLVYRADITGRYSNRLYAVEPADTALL LDNKKALKLKTTEL"
BASE COUNT 291 a 170 c 160 g 217 t
ORIGIN

Query Match 99.7%; Score 321; DB 9; Length 838;
Best Local Similarity 99.7%; Pred. No. 3.1e-83;

Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacagcaacaacacctgtagtattatcactgagtagtgagccacagcagc 60
|||||
DB 208 GACAGCAACAACCCCTTGATTTATTCACCTTGAGTAGGCCCCACACAGTCAAGC 267
QY 61 tttaagaagaagttgtctgaataaagaatccagaatctgacagagttgtcct 120
|||||
DB 268 TTTAAGAAGAAGTTGTGCTAAAATAAGAAATCCAGAAATGGCAGAGATTGTCT 327
QY 121 cctcaatctgtttatgaaacaactgacaacacctctctcctgtagtgagcc 180
|||||
DB 328 CCTCAATCTGCTTATTAAGAACCACTGACAAACCTTTCTCCTGATGCGCAGATATGCC 387
QY 181 cagagattatgttctgtagccatctctgacagtagagccgatatcaactggaagata 240
|||||
DB 388 CAGGATTATGTTGTTGACCCATCTGACAGATTAGCCGATATTCACCTGGAAGATATTC 447
QY 241 aaangctctatgctgaagaacctgagatacagctctgtctgtacacatagaagaa 300
|||||
DB 448 AAATGCTCTTATGCTTACGACACCTGAGATACAGCTCTGTGCTTGACAAATGAAGAA 507
QY 301 agctcgaagttgctgaagact 322
|||||
DB 508 AGCTCTCAAGTTGCTGAAGACT 529

RESULT 2
LOCUS ARI23828 875 bp DNA PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6171816.
ACCESSION ARI23828
VERSION ARI23828.1 GI:14109189
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 875)
AUTHORS Yu, G., Dillon, P.J., Ebner, R. and Endress, G.A.

TITLE Human XAG-1 polynucleotides and polypeptides
JOURNAL Patent: US 6171816-A 1 09-JAN-2001;
FEATURES Location/Qualifiers
source 1.875
/organism="unknown"
BASE COUNT 283 a 189 c 180 g 223 t
ORIGIN

Query Match 99.7%; Score 321; DB 6; Length 875;
Best Local Similarity 99.7%; Pred. No. 3.1e-83;

Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gacagcaacaacacctgtagtattatcactgagtagtgagccacagcagc 60
|||||
DB 268 GACAGCAACAACCCCTTGATTTATTCACCTTGAGTAGGCCCCACACAGTCAAGC 327
QY 61 tttaagaagaagttgtctgaataaagaatccagaatctgacagagttgtcct 120
|||||
DB 328 TTTAAGAAGAAGTTGTGCTAAAATAAGAAATCCAGAAATGGCAGAGATTGTCT 387
QY 121 cctcaatctgtttatgaaacaactgacaacacctctctcctgtagtgagcc 180
|||||
DB 388 CCTCAATCTGCTTATTAAGAACCACTGACAAACCTTTCTCCTGATGCGCAGATATGCC 447
QY 181 cagagattatgttctgtagccatctctgacagtagagccgatatcaactggaagata 240
|||||
DB 448 CAGGATTATGTTGTTGACCCATCTGACAGATTAGCCGATATTCACCTGGAAGATATTC 507
QY 241 aaangctctatgctgaagaacctgagatacagctctgtctgtacacatagaagaa 300
|||||
DB 508 AAATGCTCTTATGCTTACGACACCTGAGATACAGCTCTGTGCTTGACAAATGAAGAA 567
QY 301 agctcgaagttgctgaagact 322
|||||
DB 568 AGCTCTCAAGTTGCTGAAGACT 589

RESULT 3
LOCUS AF088867 956 bp mRNA PRI 02-JAN-2000
DEFINITION Homo sapiens putative secreted protein XAG mRNA, complete cds.
ACCESSION AF088867
VERSION AF088867.1 GI:6652811
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 956)
AUTHORS Zhang, J.S. and Smith, D.I.
JOURNAL Unpublished
TITLE Human homolog of XAG is differentially expressed in tumors
REFERENCE 2 (bases 1 to 956)
AUTHORS Zhang, J.S. and Smith, D.I.
JOURNAL Direct Submission
SUBMITTED (30-AUG-1998) Pathology and Lab. Medicine, Mayo Clinic, 200 SW 1st St., Rochester, MN 55905, USA
FEATURES Location/Qualifiers
source 1.956
/organism="Homo sapiens"
/db_xref="taxon:9606"
148.675
/codon_start=1
/product="putative secreted protein XAG"
/protein_id="AAE22484.1"
/db_xref="GI:6652812"
/translation="MEKIPVSAFLILVALSYTLARDFTYKPGAKKDTKSRPKLPOTL SRGWGQILWQTQYEEALYKSKTSNKPIMIIHLDECPSQALKKFAENKEIOKLAE OFVLNLVYETTDKHLSPDGQYVPRIMFVDPSTLVYRADITGRYSNRLYAVEPADTALL LDNKKALKLKTTEL"

BASE COUNT 310 a 202 c 202 g 242 t
ORIGIN

Query Match 99.7%; Score 321; DB 9; Length 956;
Best Local Similarity 99.7%; Pred. No. 3,1e-83;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacagcaacaaccccttgatgatcattatcattcattgagtgccacacagtcacg 60
DB 345 GACAGCAACAACCCCTTGATGATTAATTCATCAGTGGATGAGCCACACAGCAAGC 404

QY 61 tttaagaagaagttgttcgtgaaataaagaatccagaatctggcagagcagttgtcct 120
DB 405 TTTAAAGAAAGTGTTCCTGAAAATAAGAAATCGAGAAATGGCAGAGCAGTTGTCT 464

QY 121 cctcaatcgttttaataaactgcaaacacccttcctcattatggcagatgtccc 180
DB 465 CCTCAATCTGTTATGAAACACTGACAAACACCTTCTCTGATGAGCCAGTATGTCC 524

QY 181 cagattatgtttgttaccatctctgacagttagagccgatalcaactggaagatattc 240
DB 525 CAGGATTATGTTGTTGACCATCTCGACAGTTAGAGCCGATATACACTGGAAGTATTC 584

QY 241 aaangctctctatgctcagacgaacctgcagatacagctctgttgccttgacaacatgaaga 300
DB 585 AAACGCTCTATGCTTACGAACTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAA 644

QY 301 agcttcaagttgctgaagact 322
DB 645 AGCTCTCAAGTTGCTGAAGACT 666

RESULT 4
AX011612 1020 bp DNA PAT 06-SEP-2000
LOCUS Sequence 8 from Patent WO9955858.
DEFINITION AX011612
ACCESSION AX011612 GI:9998136
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1020)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences obtained from pancreas tumor tissue
JOURNAL Patent: WO 9955858-A 8 04-NOV-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
source 1..1020
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 341 a 214 c 218 g 247 t
ORIGIN

Query Match 99.7%; Score 321; DB 6; Length 1020;
Best Local Similarity 99.7%; Pred. No. 3,1e-83;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 181 cagattatgtttgttgcacccatctctgacagttagagccgatatcactggaagatattc 240
DB 527 CAGGATTATGTTGTTGACCATCTCTGACAGTTAGAGCCGATATACACTGGAAGTATTC 586

QY 241 aaangctctctatgctcagacgaacctgcagatacagctctgttgccttgacaacatgaaga 300
DB 587 AAACGCTCTATGCTTACGAACTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAA 646

QY 301 agcttcaagttgctgaagact 322
DB 647 AGCTCTCAAGTTGCTGAAGACT 668

RESULT 5
AX015056 1020 bp DNA PAT 07-SEP-2000
LOCUS Sequence 265 from Patent WO9953040.
DEFINITION AX015056
ACCESSION AX015056 GI:10041195
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1020)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences from ovarian tumour tissue
JOURNAL Patent: WO 9953040-A 265 21-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
source 1..1020
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 341 a 214 c 218 g 247 t
ORIGIN

Query Match 99.7%; Score 321; DB 6; Length 1020;
Best Local Similarity 99.7%; Pred. No. 3,1e-83;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacagcaacaaccccttgatgatcattatcattcattgagtgccacacagtcacg 60
DB 347 GACAGCAACAACCCCTTGATGATTAATTCATCAGTGGATGAGCCACACAGCAAGC 406

QY 61 tttaagaagaagttgttcgtgaaataaagaatccagaatctggcagagcagttgtcct 120
DB 407 TTTAAAGAAAGTGTTCCTGAAAATAAGAAATCCAGAAATGGCAGAGCAGTTGTCT 466

QY 121 cctcaatcgttttaataaactgcaaacacccttcctcattatggcagatgtccc 180
DB 467 CCTCAATCTGTTATGAAACACTGACAAACACCTTCTCTGATGAGCCAGTATGTCC 526

QY 181 cagattatgtttgttgcacccatctctgacagttagagccgatalcaactggaagatattc 240
DB 527 CAGGATTATGTTGTTGACCATCTCGACAGTTAGAGCCGATATACACTGGAAGTATTC 586

QY 241 aaangctctctatgctcagacgaacctgcagatacagctctgttgccttgacaacatgaaga 300
DB 587 AAACGCTCTATGCTTACGAACTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAA 646

QY 301 agcttcaagttgctgaagact 322
DB 647 AGCTCTCAAGTTGCTGAAGACT 668

RESULT 6
AF038451 1077 bp mRNA PRI 03-DEC-1998
LOCUS

DEFINITION Homo sapiens secreted cement gland protein XAG-2 homolog (hag-2/R)
ACCESSION AF038451
VERSION AF038451.1 GI:3779225
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1077)
AUTHORS Thompson, D.A. and Weigel, R.J.
TITLE hag-2, the human homologue of the Xenopus laevis cement gland gene
XAG-2, is coexpressed with estrogen receptor in breast cancer cell
lines
JOURNAL Biochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)
MEDLINE 99009231
REFERENCE 2 (bases 1 to 1077)
AUTHORS Thompson, D.A.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1997) Surgery, Stanford University, 1201 Welch
Road, MSLS Building, Room P228, Stanford, CA 94305, USA
FEATURES
source
1.1077
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="breast"
/cell_line="MCF7"
/note="Isolated from estrogen receptor-positive breast
adenocarcinoma cell line"
1.1077
/gene="hag-2/R"
59..586
/note="similar to Homo sapiens protein hag-2/C deposited
under GenBank Accession Number AF007791, and to Xenopus
laevis cement gland-specific proteins np77 and XAG-2
deposited under GenBank Accession Numbers U82110 and
AF025474, respectively"
/codon_start=1
/product="secreted cement gland protein XAG-2 homolog"
/protein_id="AAC82614.1"
/db_xref="GI:3779226"
/translation="MEKIPVSAFLLLVALSTLARDTVKPGAKKDKDSRPLKPLDTL
SRGWDOLMTQTEALYKSKSNKPLMIHLEDCPSQALKVFAKNKIKLAE
DFVLNLVETDKHLSPDQYVPRIMEFVPSLTVRADITGRVSNRLYAYEPADVAL
LDNMKKALKLKTTEL"
BASE COUNT 367 a 231 c 210 g 269 t
ORIGIN
Query Match 99.7%; Score 321; DB 9; Length 1077;
Best Local Similarity 99.7%; Pred. No. 3.2e-83;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gacaaacaacacccctgatgatattcaatccttgatgatgagccacacagtcgaagc 60
|||||
DB 256 GACAAACAACAAACCCCTGATGATATCATCTTGATGATGAGCCACACAGTCAAGC 315
QY 61 tttaagaagaagtgctgcgaataaagaatccaagaattgacgagagattgtcct 120
|||||
DB 316 TTTAAAGAAAGTGTGCTGAAATAAAGAAATCCAGAAATTCGACAGAGTTTGTCTT 375
QY 121 cctcaatcgtgtttatgaacaacacacaccccttcctcgcgaatggcgaatgtccc 180
|||||
DB 376 CCTCATCTGGTTTATGAACAACACACCTTTCTCCTGATGGCCAGATATGCCC 435
QY 181 caggaattatgttcttgaccatctcgcagttagagccgataactggaagataatc 240
|||||
DB 436 CAGGATTATGTTGTTGACCACTCTGACAGTTAGAGCCGATATACACTGGAAGATATTC 495
QY 241 aaangctcctatgcttaagaacctgcagataagctctgtcttgacaaactgaagaa 300
|||||
DB 496 AATCGTCTATGCTTACGAAACCTGCAGATACAGCTGTGCTTGCACAAACATGAAGAA 555

QY 301 agctcgaagttgctgaagact 322
|||||
DB 556 AGCTTCAAGTTGCTGAAGACT 577
RESULT 7
AX156194/c 529 bp DNA PAT 22-JUN-2001
LOCUS AX156194/c
DEFINITION Sequence 51 from Patent WO0140269.
ACCESSION AX156194
VERSION AX156194.1 GI:14537195
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 529)
AUTHORS Dillon, D.C., Day, C.H., Jiang, Y., Houghton, R.L., Mitcham, J.L. and
Wang, A.
TITLE Compositions and methods for therapy and diagnosis of breast cancer
JOURNAL Patent: WO 0140269-A 51 07-JUN-2001;
CORIXA CORPORATION (US)
FEATURES
source
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 147 a 104 c 102 g 176 t
ORIGIN
Query Match 96.3%; Score 310; DB 6; Length 529;
Best Local Similarity 99.4%; Pred. No. 4.8e-80;
Matches 321; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 gacaaacaacaccccttgatga-ttatcattcacttgatgatgagtcacacagtcgaag 59
|||||
DB 507 GACAAACAACAAACCCCTGATGATATTCATCTTGATGATGAGCCACACAGTCAAG 448
QY 60 cttaagaagaagtgctgcgaataaagaatccaagaattgacgagagattgtcct 119
|||||
DB 447 CTTTAAAGAAAGTGTGCTGGAATAATTAAGAAATCCAGAAATTCGACAGACATTTGTCC 388
QY 120 tctcaatcgtgtttatgaacaacacacacaccccttcctcctgaatggcgaatgtccc 179
|||||
DB 387 TCTCATATCTGGTTTATGAACAACAGTACAAACACCTTCTCTGATGGCCATATGTC 328
QY 180 ccaagattatgttcttgaccatctcgcagttagagccgatatcaactggaagatat 229
|||||
DB 327 CCAGGATTATGTTGTTGACCCATCTCTGACAGTTAGAGCCGATATCAGTGAAGATATT 268
QY 240 caaangctcctatgcttgaagaacctgcagataagctctgtcttgacaaactgaagaa 299
|||||
DB 267 CAACCGTCTTATGCTTACGAAACCTGCAGATACAGCTGTGCTTGCACACATGAAGA 208
QY 300 aagctcgaagttgcgaagact 322
|||||
DB 207 AACCTCAAGTTGCTGAAGACT 185
RESULT 8
AX067336 778 bp DNA PAT 24-JAN-2001
LOCUS AX067336
DEFINITION Sequence 40 from Patent WO0078960.
ACCESSION AX067336
VERSION AX067336.1 GI:12544960
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 778)
AUTHORS Yudin, J. and Mitcham, J.L.

TITLE Compositions and methods for the therapy and diagnosis of breast cancer
JOURNAL Patent: WO 0078960-A 40 28-DEC-2000;
CORIXA CORPORATION (US)
FEATURES Location/Qualifiers
Source 1..778
/organism="Homo sapiens"
/db_xref="taxon:9606"
misc_feature 1..778
/note="n = A,T,C or G"
BASE COUNT 245 a 175 c 158 g 193 t 7 others
ORIGIN

Query Match 96.0%; Score 309; DB 6; Length 778;
Best Local Similarity 99.4%; Pred. No. 9.8e-80;
Matches 320; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 gacaaagcaaacacccctgatgatattatcatcacttgatgagtgccacacagtcacgc 60
|||||
Db 256 GACAGCAGCAACACCCCTTGATGATTATTCATCCTTGATGAGTCCACACAGTCAAGC 315
|||||

QY 61 tttaagaaagtgttctgcaaaaataaagaatccagaattggcagagcgattgtcct 120
|||||
Db 316 TTTAAAGAAAGTGTTCGCTGAATAAAGAAATCCAGAAATTGGCAGAGCATTTGTCTCT 375
|||||

QY 121 cctcaatctgttttaagaacactgacacacacttctctgtagcgagatgtcc 180
|||||
Db 376 CCTCAATCTGTTTATGAAACACTGACAAACCTTCTCTGATGGCCAGTATGTCC 435
|||||

QY 181 caggaattatgttctgacccatctctacagcttagagcagatcactggaagatctc 240
|||||
Db 436 CAGATTATGTTTGTGACCCATCTCTACAGTTAGAGCCATATCATTGGAGATATTC 495
|||||

QY 241 aaancgtctctatgcttaagaacctgcagatcacagctctgttgcctgaacaatgaaga 300
|||||
Db 496 AAATCGTCTATGCTTACGAACTGCAGAT-CAGCTCTGTTCCTGAACAATGAAAGAA 554
|||||

QY 301 agcttcaagtgtctgaagact 322
|||||
Db 555 AGCTCTCAGTTGCTGAAGACT 576
|||||

RESULT 9
LOCUS ARI23857 386 bp DNA PAT 16-MAY-2001
DEFINITION Sequence 35 from patent US 6171816.
ACCESSION ARI23857
VERSION ARI23857.1 GI:14109218
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 386)
AUTHORS Yu,G., Dillon,P.J., Ehner,R. and Endress,G.A.
TITLE Human XAG-1 polynucleotides and polypeptides
JOURNAL Patent: US 6171816-A 35 09-JAN-2001;
FEATURES Location/Qualifiers
Source 1..386
/organism="unknown"
BASE COUNT 122 a 95 c 74 g 95 t
ORIGIN

Query Match 85.4%; Score 275; DB 6; Length 386;
Best Local Similarity 99.6%; Pred. No. 7.9e-70;
Matches 275; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaaagcaaacacccctgatgatattatcatcacttgatgagtgccacacagtcacgc 60
|||||
Db 111 GACAGCAGCAACACCCCTTGATGATTATTCATCCTTGATGAGTCCACACAGTCAAGC 170
|||||

QY 61 tttaagaaagtgttctgcaaaaataaagaatccagaattggcagagcgattgtcct 120
|||||

|||||
Db 171 TTTAAAGAAAGTGTTCGCTGAATAAAGAAATCCAGAAATTGGCAGAGCATTTGTCTCT 230
|||||

QY 121 cctcaatctgttttaagaacactgacacacacttctctgtagcgagatgtcc 180
|||||
Db 231 CCTCAATCTGTTTATGAAACACTGACAAACCTTCTCTGATGGCCAGTATGTCC 290
|||||

QY 181 caggaattatgttctgacccatctctacagcttagagcagatcactggaagatctc 240
|||||
Db 291 CAGATTATGTTTGTGACCCATCTCTACAGTTAGAGCCGATATCATTGGAGATATTC 350
|||||

QY 241 aaancgtctctatgcttaagaacctgcagatcacagctctgttgcctgaacaatgaaga 276
|||||
Db 351 AAATCGTCTATGCTTACGAACTGCAGATTCAGATCAAGC 386
|||||

RESULT 10
LOCUS AXI92465 401 bp DNA PAT 15-AUG-2001
DEFINITION Sequence 32 from patent WO0149716.
ACCESSION AXI92465
VERSION AXI92465.1 GI:15210429
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 401)
AUTHORS Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,
Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.
TITLE Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use
JOURNAL Patent: WO 0149716-A 32 12-JUL-2001;
CORIXA CORPORATION (US)
FEATURES Location/Qualifiers
Source 1..401
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 128 a 101 c 77 g 95 t
ORIGIN

Query Match 81.8%; Score 263.4; DB 6; Length 401;
Best Local Similarity 98.9%; Pred. No. 1.9e-66;
Matches 275; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 gacaaagcaaacacccctgatgatattatcatcacttgatgagtgccacacagtcacgc 60
|||||
Db 124 GACAGCAGCAACACCCCTTGATGATTATTCATCCTTGATGAGTCCACACAGTCAAGC 183
|||||

QY 61 tttaagaaagtgttctgcaaaaataaagaatccagaattggcagagcgattgtcct 120
|||||
Db 184 TTTAAAGAAAGTGTTCGCTGAATAAAGAAATCCAGAAATTGGCAGAGCATTTGTCTCT 243
|||||

QY 121 cctcaatctgttttaagaacactgacacacacttctctgtagcgagatgtcc 180
|||||
Db 244 CCTCAATCTGTTTATGAAACACTGACAAACCTTCTCTGATGGCCAGTATGTCC 303
|||||

QY 181 caggaattatgttctgacccatctctacagcttagag-cagatcactggaagatct 239
|||||
Db 304 CAGATTATGTTTGTGACCCATCTCTACAGTTAGAGCCCGATATCATTGGAGATATTC 363
|||||

QY 240 caaancgtctctatgcttaagaacctgcagatcacagctctgttgcctgaacaatgaaga 277
|||||
Db 364 CAACCGTCTATGCTTACGAACTGCAGATCAAGTCAAGT 401
|||||

RESULT 11
LOCUS AB016592 758 bp mRNA ROD 10-NOV-1999
DEFINITION Mus musculus mRNA for G0B-4, complete cds.
ACCESSION AB016592
VERSION AB016592.1 GI:3395165

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

GOB-4.
Mus musculus (strain:Balb/c) adult intestine goblet cell cDNA to mRNA, clone:GOB-4.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
Komiya,T., Tanigawa,Y. and Hirohashi,S.
Cloning of the gene gob-4, which is expressed in intestinal goblet cells in mice
Biochim. Biophys. Acta 1444 (3), 434-438 (1999)
99196718
2 (bases 1 to 758)
Komiya,T. and Hirohashi,S.
Direct Submission
Submitted (29-JUL-1998) to the DDBJ/EMBL/Genbank databases. Tohru Komiya, ERATO, JST, Genome Asymmetry Group, Doi Biosymmetry Project, Tokodai 5-9-9, Tsukuba, Ibaragi 300-2635, Japan
(E-mail:tkome@bio.eratec.net.co.jp, Tel:81-298-48-1515, Fax:81 298-47-8901)

FEATURES
source
location/Qualifiers
1..758
/organism="Mus musculus"
/strain="Balb/c"
/db_xref="taxon:10090"
/cell_type="goblet cell"
/clone="Gob-4"
/dev_stage="adult"
/tissue_type="intestine"
52..579
/gene="gob-4"
52..579
/gene="gob-4"
/codon_start=1
/product="GOB-4"
/protein_id="BA332044.1"
/db_xref="GI:3395166"
/translation="MEKFSVSAILLVAISGLAKDTYVSGAKKDPKDSRPRLPQTL SRMGDQLWTQTYEALYRSKTSNRPILVTHHLDPCPSOALKVFAHKEIOKLAE QEVLLNLVYETDVKHLSPDGQYVPRIVFVPSLTVRADITGRYSNRLYAVEPSDIAL YDNMKKALKLKTTEL"
758
polyA_site
/note="18 a nucleotides"
BASE COUNT 234 a 185 c 165 g 174 t
ORIGIN

Query Match 81.5%; Score 262.4; DB 10; Length 758;
Best Local Similarity 88.5%; Pred. No. 4.1e-66;
Matches 284; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 gacaagcaaacacccctgatgatattcatcacttgatgatgagtgccacacagtcgaagc 60
|||||
DB 249 GACACAGCAACAGACCCCTTGATGTCATTCATCTTGAGAGAAATGCCACACATCAAGC 308
|||||
QY 61 tttaagaaagtgcttgctgaataaagaatccagaatctggcagagcagttgtcct 120
|||||
DB 309 CTTAAAGAAAGTGTGCTGACATTAAGAAATCCAGAAATGGCAGAGAGTTGTTCT 368
|||||
QY 121 cctcaatctggtttatgaacaactgacaacaccttctccctgattggcagatattgtccc 180
|||||
DB 369 CCTCAACCTGTGTATGAACAACCGCAACGACCTTTCTCTGATGGCAGTAGCGTCCC 428
|||||
QY 181 caagatattgtttgaccacatctcgacagttagacgagatatacactgaaatattc 240
|||||
DB 429 CACAATTTGTTTGTAGACACCATCCCTGACGGTGAGGCGACATCTCTGGACATACCTC 488
|||||
QY 241 aaanagctctatagcttaagaaactgacagaataagctctgttgcttgacacatagaaga 300
|||||
DB 489 AAACCGGCTCTAGCGTTATGACCTCTGTACACAGCTTTGTTGACGACAACTGAAGAA 548
|||||
QY 301 agctctcaagttgctctgaagac 321
|||||

Db 549 AGCTCTCAAGCTGCTAAAGAC 569

RESULT 12
AF044262
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

AF044262 781 bp mRNA ROD 11-NOV-1998
Mus musculus secreted cement gland protein XAG-2 homolog (mAg-2)
mRNA, complete cds.
AF044262
AF044262.1 GI:3779231
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 781)
Thompson,D.A. and Weigel,R.J.
hAg-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines
Biochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)
9909231
2 (bases 1 to 781)
Thompson,D.A.
Direct Submission
Submitted (22-JAN-1998) Surgery, Stanford University, MSLS Building, Room P228, 1201 Welch Road, Stanford, CA 94305, USA

FEATURES
source
location/Qualifiers
1..781
/organism="Mus musculus"
/strain="Swiss Webster"
/db_xref="taxon:10090"
/dev_stage="17-day embryo"
1..781
/gene="mAg-2"
53..580
/gene="mAg-2"
/note="similar to Homo sapiens proteins hAg-2/R deposited under GenBank Accession Number AF038451 and hAg-2/C deposited under GenBank Accession Number AF007791, and to Xenopus laevis cement-gland proteins m97 deposited under GenBank Accession Number U82110 and XAG-2 deposited under GenBank Accession Number AF025474"
/codon_start=1
/product="secreted cement gland protein XAG-2 homolog"
/protein_id="AAC72705.1"
/db_xref="GI:3779232"
/translation="MEKFSVSAILLVAISGLAKDTYVSGAKKDPKDSRPRLPQTL SRMGDQLWTQTYEALYRSKTSNRPILVTHHLDPCPSOALKVFAHKEIOKLAE QEVLLNLVYETDVKHLSPDGQYVPRIVFVPSLTVRADITGRYSNRLYAVEPSDIAL YDNMKKALKLKTTEL"
781

BASE COUNT 256 a 185 c 166 g 174 t
ORIGIN

Query Match 81.5%; Score 262.4; DB 10; Length 781;
Best Local Similarity 88.5%; Pred. No. 4.1e-66;
Matches 284; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 gacaagcaaacacccctgatgatattcatcacttgatgatgagtgccacacagtcgaagc 60
|||||
DB 250 GACACAGCAACAGACCCCTTGATGTCATTCATCTTGAGAGAAATGCCACACATCAAGC 309
|||||
QY 61 tttaagaaagtgcttgctgaataaagaatccagaatctggcagagcagttgtcct 120
|||||
DB 310 CTTAAAGAAAGTGTGCTGACATTAAGAAATCCAGAAATGGCAGAGAGTTGTTCT 369
|||||
QY 121 cctcaatctggtttatgaacaactgacaacaccttctccctgattggcagatattgtccc 180
|||||
DB 370 CCTCAACCTGTGTATGAACAACGCAACGACCTTTCTCTGATGGCAGTAGCGTCCC 429
|||||
QY 181 caagatattgtttgaccacatctcgacagttagacgagatatacactgaaatattc 240
|||||

Db	430	CAGAAATTGGTTTGAAGAACCCATCCCTGACGGTGAAGGGGACAGACATTCACATCGACGATATAC	489
Oy	241	aaancgtctctatgcttctcagcaactcgacagatacaagctctgttgccttgatgacacatgaaga	300
Db	490	AAACGGGCTTACGGCTTATGAAACCTTCTGACACAGCTTGTGTACAGACATGTAAGA	549
Oy	301	agctctcaattgctcgaagac	321
Db	550	AGCTCTCAAGCTGCTAAAGAC	570
RESULT	13		
LOCUS	AX192616		
DEFINITION	Sequence 183 from Patent WO0149716.	PAT	15-AUG-2001
ACCESSION	AX192616		
VERSION	AX192616.1		
KEYWORDS	GI:15210579		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	EnxarVota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	Xu,J., Lodes,M.J., Seisist,H., Benson,D.R., Meagher,M.J., Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.		
TITLE	Compounds for immunotherapy and diagnosis of colon cancer and methods for their use		
JOURNAL	Patent: WO 0149716-A 183 12-JUL-2001; CORIXA CORPORATION (US)		
FEATURES	location/qualifiers		
Source	1..409		
BASE COUNT	131 a 104 c 80 g 94 t		
ORIGIN	/db_xref="taxon:9606"		
Query Match	77.3%; Score 249; DB 6; Length 409;		
Best Local Similarity	99.6%; Pred. No. 3.2e-62;		
Matches 249; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
Oy	1	gacaaagcaacaaaccttgatgatattatcattacttgatggtgcgccacaaagccaagc	60
Db	160	GACAAAGCAACAAACCCCTGATGATTAATTCATCATCTTGATGAGTGCCCCACACAGTCAAGC	219
Oy	61	tttaagaagaagtgttctcgtaaaatacaagaatccaagaatggcagaagcagttgtcct	120
Db	220	TTTAAAGAAAGCGTTTGCCTGCAAAATTAAGCAATCCAGAAATTGGCAGACAGTTGTCTCT	279
Oy	121	cctcaatcgtgttattgaacaactgacaacaacaccttctcctgtatggccaglatgtccc	180
Db	280	CCTCAATCTGGTTTATGAAACAACAGACAAACACCTTCTCCTGATGGCCAGTATGTCCC	339
Oy	181	cagagttatgtttgttgcccccacatcctcgcacagttagagccgatatcactcggagaatattc	240
Db	340	CAGGATTAATGTTGTTGACCCATCTCTGACAGTTAGAGCCGATATTCACGAAAGATATTC	399
Oy	241	aaancgtctc 250	
Db	400	AAATGCTCTC 409	
RESULT	14		
LOCUS	ARI44213	793 bp	DNA
DEFINITION	Sequence 78 from patent US 6210883.	PAT	08-AUG-2001
ACCESSION	ARI44213		
VERSION	ARI44213.1		
KEYWORDS	GI:15106080		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 793)		

FEATURES	source	location/Qualifiers	1..793
BASE COUNT	246 a	172 c	148 g
ORIGIN		211 t	16 others
Query Match	Best Local Similarity	76.6%;	Score 246.8; DB 6; Length 793;
Matches	299; Conservative	0; Mismatches	16; Indels 5; Gaps 4;
Y	1	gacaaagcaacaacccctgatgatattcatcacttgatgagtcgcccacagtcacg	60
Db	253	GACAAAGCAACAACCCCTTGATGATATTATTCATACACTGGATGATGCCACACAGTCNNAGC	312
Y	61	tttaagaagaatggttgcctgaaataaagaatccagaatgtgcagagcagttgtcct	120
Db	313	TTTAAAGAAAGGTGTGCTGAAAAATTAAGAAATCCAGAAATTTGGCAGAGCAGATTGTCTCT	372
Y	121	cctcaatctgttttaatgaacaact-gacaaacaccccttctcctgattgagcagatagtc	180
Db	373	CCTCAATCTGTTTATGTAACACTGACAAACACCTTCTCTGATGGCCAGTATGT-CC	431
Y	181	cagagatlatggtttgtgacccatctctacagtgtaagcagatcacttggaagatattc	240
Db	432	CAGGATTAATGTTGTTGGTGAACCCATCTCTCAGAGTTAAACCCATATCCGGAGAGATATTC	491
Y	241	aaanccgtctctatgcttaagaaacctgcagatacagctctgttgccttgacacatgaagaa	300
Db	492	NAACCGTCTCTTAATGCTTAC-AAACGACAGATAC-SCCTGTGCTTGACAA--CATGAATA	547
Y	301	agctctcaagtgtcgtgaaga	320
Db	548	AGCTCTCAAGTTGCTMAAA	567
RESULT	15	ARI23832	
LOCUS	ARI23832	506 bp	DNA
DEFINITION	Sequence	10 from patent	US 6171816.
ACCESSION	ARI23832		
VERSION	ARI23832.1	GI:14109193	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	YU,G., DILLON,P.J., Ebner,R. and Endress,G.A.		
TITLE	Human XAG-1 polynucleotides and polypeptides		
JOURNAL	Patent: US 6171816-A-10 09-JAN-2001;		
FEATURES	Location/Qualifiers		
Source	1..506		
BASE COUNT	131 a	84 c	125 g
ORIGIN		143 t	23 others
Query Match	Best Local Similarity	63.9%;	Score 205.8; DB 6; Length 506;
Matches	229; Conservative	0; Mismatches	5; Indels 2; Gaps 2;
Y	1	gacaaagcaacaaccccttgatgatattcatcacttgatgagtcgcccacagtcacg	60
Db	16	GACAAAGCAACAACCCCTTGATGATATTATTCATCANTTGGATGATGCCACACAGTCACAGC	75
Y	61	tttaagaagaatggttgcctgaaataaagaatccagaatgtgcagagcagttgtcct	120
Db	76	TTTAAAGAAAGGTGTGCTGAAAAATTAAGAAATCCAGAAATTTGGCAGAGCAGTTGTCTCT	135
Y	121	cctcaatctgttttaatgaacaact-gacaaacaccccttctcctgattgagcagatagtc	179

```
Db 136 CCTCAATCTGTTATGAAACAACAGCAACACCTTCTCCGATGGCCAGTATGTNC 195
OY 180 ccagattatgtttgttgaccatcct-gacagtiagagccgatatcactggaag 234
Db 196 CCAGGTTATGTTGTGTGACCATNTCTGAGACAGTTAGAGCCGATATCACTGGGAG 251
```

Search completed: March 30, 2002, 12:59:18
Job time: 21440 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 09:31:11 ; Search time 5438.31 Seconds
(without alignments)
9028.073 Million cell updates/sec

Title: US-09-867-034-3

Perfect score: 4569

Sequence: 1 atgataaatcctcaacac.....tcagctgtcaataagcttag 4569

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estlda:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_estl:*
11: gb_estc2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_juv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1514.4	33.1	2933	12	AK008659 Mus muscu
2	1500.2	32.8	2915	12	AK007466 Mus muscu
3	808	17.7	2703	13	AF101616 AF101616
4	655	14.3	825	10	AU122701 AU122701
5	631	13.8	745	11	BG622206 BG622206
6	630.4	13.8	943	13	AO748956 HS_5574_A
7	629.6	13.8	810	13	AO629742 RBC1-11-4
8	620.2	13.6	768	13	AO740881 HS_2274_A
9	617.8	13.5	893	10	AU137259 AU137259
10	615.2	13.5	863	13	AO752204 HS_3571_B
11	611	13.4	653	10	AW361534 OV2-CT026
12	610	13.4	872	13	AO896432 HS_3134_A

13	608.6	13.3	860	13	AO747583	AO747583 HS_5533_A
14	606	13.3	796	13	AO636664	AO636664 RPT1-11-4
15	605.4	13.3	759	13	AO058259	AO058259 CIT-HSP-2
16	603.4	13.2	734	10	AI660957	AI660957 wf20408.X
17	603.2	13.2	756	13	AO041614	AO041614 CIT-HSP-2
18	595.6	13.0	897	13	AO746355	AO746355 HS_2278_A
19	592.8	13.0	820	13	AO746771	AO746771 HS_2276_A
20	589.4	12.9	793	13	AO542629	AO542629 RPT1-11-3
21	586.2	12.8	891	13	AO746363	AO746363 HS_2278_A
22	584.8	12.8	1393	12	AF116619	AF116619 Homo. Sap1
23	582.6	12.8	828	13	AO780863	AO780863 HS_3113_B
24	580	12.7	689	13	BI6435	BI6435 342D19.TPB
25	578.8	12.7	712	11	BG699667	BG699667 60261514
26	578.6	12.7	863	13	AO746062	AO746062 HS_2277_A
27	578.6	12.7	894	13	AO739350	AO739350 HS_5482_B
28	572.2	12.5	762	13	AO200047	AO200047 RPT11-59
29	572.2	12.5	763	13	AO039481	AO039481 CIT-HSP-2
30	569	12.5	925	11	BG403000	BG403000 602418809
31	567.8	12.4	598	11	BE927204	BE927204 RCI-CT024
32	567	12.4	766	13	AO745976	AO745976 HS_2275_A
33	566.6	12.4	633	10	AW753451	AW753451 OV2-CT026
34	565	12.4	835	13	AO780914	AO780914 HS_3113_B
35	564.8	12.4	734	13	AO262912	AO262912 CIT-B1-E1
36	563.6	12.3	712	13	AO055831	AO055831 CIT-HSP-2
37	562	12.3	721	13	AO395312	AO395312 CIT-B1-E1
38	560.8	12.3	710	13	AO077037	AO077037 CIT-HSP-2
39	560.2	12.3	730	13	B59966	B59966 CIT-HSP-347
40	558.6	12.2	723	13	AO383358	AO383358 RPT11-13
41	558.6	12.2	842	13	AO747087	AO747087 HS_5538_A
42	557.4	12.2	709	13	B92460	B92460 CIT-HSP-217
43	555.4	12.1	625	10	AW361532	AW361532 OV2-CT026
44	554.4	12.1	782	13	AO269601	AO269601 HS_2036_B
45	552.6	12.1	712	13	AO047421	AO047421 CIT-HSP-2

ALIGNMENTS

RESULT 1
AK008659
LOCUS
DEFINITION
Mus musculus adult male stomach cDNA, RIKEN full-length enriched library, clone:2210006G11, full insert sequence.
ACCESSION
AK008659
VERSION
AK008659.1 GI:12842987
KEYWORDS
CAP trapper.
SOURCE
Mus musculus (strain:C57BL/6J) adult male stomach cDNA to mRNA, clone:11b:RIKEN full-length enriched mouse cDNA library
clone:2210006G11.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 2933)
Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
METHODS
Methods in enzymology. 303, 19-44 (1999)
PUBMED
9927923
REFERENCE
2 (bases 1 to 2933)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalisation and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
METHODS
Genome research. 10 (10), 1617-1630 (2000)
JOURNAL
20499374
MEDLINE
11042159
PUBMED
3 (bases 1 to 2933)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, T., Nishi, K., Kikunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishie, T., Harada, A., Yamamoto, R., Matsuno, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,

Db	968	AAGTCGGGAGACGCTGGAACGATGATCGCTTTAACCCGATAAATTCAGGAACCGCGCTT	1021
Oy	2725	ttctctgcgcgcagaaacagtttgacgtcgtgggtccctgggttggagatgltgacatlttgacagtgct	2784
Db	1028	TTCTCTGTCGACACTGTGGAGCGACGGGATCCTGGCTGGGATGGTGCCTTTACAGTCTCT	1087
Oy	2785	gcccatgtacaataatgaattccatacagataaacaagttgcagttgcacagggacacactgcgc	2844
Db	1088	GCCATATGACAAACCGAAACCTCAACAGTTTAAACAGTGGTGGTGCACAGAGATTCGTCTGATC	1147
Oy	2845	aaaagattaccttgacgaagctttagagaggagtcacatctgcaggggcttgcattcgca	2904
Db	1148	AAGCACTTACCCACAGATATTCGACGGAGGGACATCTATATGCTCTGGCCCTTCGACACCA	1207
Oy	2905	tttactgatatgtgycacaatltgcctgltttlccatgacacacagcaatltatvggagtg	2964
Db	1208	TTTTAC-----	1212
Oy	2965	cgacaagaanaatccaatctggcctctctgycctcagcttagtgattagagaanaatat	3024
Db	1213	-----AGTATTAAGAGAAAGTAT	1231
Oy	3025	ccaactgatgatctggaatttgcctgcctgcgaagatgggggaagaacaaactataatgtgg	3084
Db	1232	CCAACCTATGATGTGAATTTGCTGCTGCACCGATGGGGAGGACCAACCACTTAGCGAGC	1291
Oy	3085	tgctttaacgaggtccaacaacaagtgtgtccatcatccacacagtcgcttggggccctct	3144
Db	1292	TGCTTTGACCCGGTGAGGCGACGGGGGCCATCATACAGTGTGGCCTGGGACGGCT	1351
Oy	3145	gcagctcaagaactagagagagctgtgccaaaatgacagagaggttaacagacatgtctca	3204
Db	1352	GCCCTTAAGAGCTTGAGCGAGCTGCTCAAAATACAGAGGCGCTCGACAGCATCTCTCG	1411
Oy	3205	gataaagttaagaacaatlggcctccattgattgcttggggcccttcaatcaggaaatgta	3264
Db	1412	GATGAGGTTTCAGAACAAATGGTCTTTGTTGATGCTTTTGCGAGCACTCCTCCAGAAATCG	1471
Oy	3265	gctgtcctcagcgtccatccaccagcttgagagtaagaggttaacctccagaacacgcag	3324
Db	1472	GGCATCGCTCAGCACTCCATTCACGCTGGAGAGCAGGGGAGTTAATCTCCAGAAATACCA	1531
Oy	3325	tgagatgaatggcacagatgatacgttggagaagcaaccgttggagaagacacttghttctatc	3384
Db	1532	TGGATGTATGGCTCAGTACGATCGTGGACAGCTCGTGGCAAGACACCTTGTTCTTATC	1591
Oy	3385	accggagcaatlgcagcctccccaactccctcttggatcccaatggcagcggagagaagaagt	3444
Db	1592	ACCTGGACAAACGCATCTCTCTACAAATTTATTTGGGATCCCAAGCGAGTGGAAATAAT	1651
Oy	3445	gagcttgtagtggacaaaacacacaaatbgccttaccctcaatcccaatcccgacatgtctag	3504
Db	1652	GGTTTATTACTAAGACAAACCACTAAGTGtGGCTACTCTCAAGTCCAGCGACGGCTAAG	1711
Oy	3505	gttggacacttggaaatacagltctgcaagaagctcacaacacttgcacctgactgtgtcacg	3564
Db	1712	GTTGGCTTTTGGAAATACACATTTCAAGGAGGCTCACAGACTCTCACCTTGCATGTCAAC	1771
Oy	3565	tcccggtgcgtccaatgttaccctgcctccactccaatgactgacttccaaaacgaacaagagc	3624
Db	1772	TCCCGTGCAGCAATGCTTACACGTCTCTTATTACAGTACACCCCGGTAGTGAATTAAGAC	1831
Oy	3625	accegaaatccccaacccctctgtgttagtttagtaataatcgcgaagagagcctccccca	3684
Db	1832	ACAGGGAATTTCCCGACGCCCTGTAAACAGTGTATGACAGCAATTCGCCCAAGGAACCTCGCT	1891
Oy	3685	attctcagggccagttgtcacagccctgfatgtaatcagltgaaatgtaaaaaacagttactctg	3744
Db	1892	ATTCTCAGGGGCGACGTCACACGCTTGATTTGATATCTGTGAATGAGAAAAACAGTAAACCTG	1951
Oy	3745	gaactactgataatgtgagcaagtgctgtatgctactaagaagtgcagcgtgtctactcaag	3804
Db	1952	GAATTAATCTGATTAACGGAGCAGGTGGCGATGCCACCCAAAGAAATGAAGTGTCTACTACTAAG	2011

QY	3805	tattttcaacttattgacacgaaatggttagatacaagtgtaaaagtgcgggctctgggaagga	3864
Db	2012	TTTTTTACAGCTTTTGGATGCAAAATGGTAGATACAGCCTTTAAATATATGGGCTCTGGAGAGA	2071
QY	3865	gttaacgcagcccaacacggagagtgatbaccacagagagtggaacactgtacatactgcgc	3924
Db	2072	GTCACTTACAGACAGACAGAGACAGCACCTTCGAGAACAGACCCATTGTACATATAGATGGC	2131
QY	3925	tggaattgagaaatgtaaaatlaacaatggaaatccacaagaacctgaaatlaaagaaatg	3984
Db	2132	TGGATTGGAGATGGGGAGAGTAGAATGACACCCACACAGCTCTGAACTATGTT-----AT	2185
QY	3985	gttcaacacaaagcaagatggtgttlltaagcaagaacatctcggagaggtcatattgtgcttct	4044
Db	2186	GTTCAGAGCAAGACACACTCTGCTTAGAGAGACATCTTCAGGGGGATCGTTGTGGCGACCC	2245
QY	4045	gattgccc---caaatgctcccataccgatctctctccacactggccaatatcacagacctg	4101
Db	2246	AATGTCGCCCGCAGCAGCAGCTCCCATTCCTGACTCTTTCCACCCCTGTCAATACACTGACCTG	2305
QY	4102	aagcggaataatcacaagggggagctcatcattaaatctgacttggagacagctctcgggagat	4161
Db	2306	AAGGCACACATCCAAAGGSCAGAACCTGGTAGATGTGACGTGAGAGCGGCTCGGGAGATGAG	2365
QY	4162	tatgaccttgaaacagcctccacaagatatacatcttgaaataagtaagatattcttgatctc	4221
Db	2366	TACGACCAACGGGAGAGCTTCACATATACATCTCCGATAGAGCACACATATCTTTGATCTC	2425
QY	4222	aagagacaagttlcaatgaaatcctctcaagttagatactactgtctcatcccaagaagagcc	4281
Db	2426	AGGACACACTTCACAACTCACTCCAACTGAAACACTCCGCTCTTATCCCAAGAGGCC	2485
QY	4282	aactctgaggaagtccttttglttaaaccgagaaacattactttgaaaaatggcacagat	4341
Db	2486	AGCTCTGAGAAACTTTTGGATTGTGAACCTGGAGGCAACACTTTTGGAAATGGCACAGAT	2545
QY	4342	ctttccattgctatctcagcgtcgtttgataaagttgatatcttgaaatcagaataatccaatt	4401
Db	2546	ATCTTCATTTGCTATCCACAGCGCTGTGGATGATGACCAATCTGAAATCAGAAATCTCCACATT	2605
QY	4402	gcaacagatctctgtttatctccccaacagactcgcgcaagagacacctgactctgataa	4461
Db	2606	GCAGCGGTCTGTGTTCATCCCGCTCAG-----GAGCGCGCCATTCCCGAAGAC	2656
QY	4462	acgtctgctcctgttccataatcatatacaagaacacatctcctggcatcaattta	4521
Db	2657	TCAACTCCCTCTGCTCCGATCATGAGATCAACAGCACACTTCCTGGCATCCACAGTCTG	2716
QY	4522	aaaattatgtgaagtgatagaggaagactgagctgcatatagacctg	4569
Db	2717	AGGATTAATGTGAAATGTGGCTAGGGGAATGTGAGGTGACACTAGAGTTTG	2764

ORGANISM

REFERENCE

MEDLINE

PUBMED
10349636
REFERENCE
2 (bases 1 to 2915)
AUTHORS
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome research. 10 (10), 1617-1630 (2000)

JOURNAL
MEDLINE
20499374
PUBMED
11042159
3 (bases 1 to 2915)
REFERENCE
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanuki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipettor sequencer Genome research. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
20530913
PUBMED
11076861
4 (bases 1 to 2915)
REFERENCE
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
5 (bases 1 to 2915)
AUTHORS
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arkawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirokawa,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Katoh,H., Kawaji,J., Koike,Y., Kono,H., Kouda,K., Nomura,K., Koyama,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sobabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and genome science laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGCAGACGAGACATCCAGACTCTTTTTCCTTTTTTTTN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence15' GAGCAGACGAGCGGCCGCAATTATCTGCAGTATTAAATTAATACC6CCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES
SOURCE
Location/Qualifiers
1..2915
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1346342"
/db_xref="MGD:MGI:1893339"
/clone="1810012P03"
/sex="male"
/tissue_type="pancreas"
/clone_lib="RIKEN full-length enriched mouse cDNA library"

	BASE COUNT	838 a	720 c	695 g	662 t	/dev.stripe="10 day old"
	ORIGIN					
	Query Match	32.8%	Score 1500.2:	DB 12:	Length 2915;	
	Best Local Similarity	74.9%:	Pred. No. 0:			
	Matches 2013:	Conservative	0:	Mismatches	558:	Indels 118; Gaps 6;
OY	1888	aaggaacatgtagaccagaacalcctcgtatctgcgttcgaagtacagaaagcaattttat	1947			
Db	188	AAGGACATGGTACTTCAGGCCCTCCCATACCGTTTGAAAGCTACAGAAAAGATTTTAC	247			
OY	1948	tccaataatgttgcattttgatctccctyaaaacatygaagacaagaagctgaactatytgaa	2007			
Db	248	TTCAAAAATGTGGCATTGTTGAATTCACAGAGCTGGAGAAGCAAAACCTGAATATAACGAGG	307			
OY	2008	cnaaaccttgaaaccttaaaaaaatgtctatitcttgtgtgttgtagtctaacctcccaagt	2067			
Db	308	CCAATACCTTGAACCTCTTAANAACCGTATGTCCTTGATATACAACACCACCTCTTAGGC	367			
OY	2068	aatga.tgaacccctaactcagtcagat.tggcgcaactigtgagagaaggtygaaggatccac	2127			
Db	368	AATGATGAGCCCTACACCGAATATPAGACATGSTGAGAAAAGGGATACAGATTAC	427			
OY	2128	ctcactcctgatattcaattgcagsgaaaaaagtlagctgaatatlygaaccaagaagtaygca	2187			
Db	428	CTGACTCTCTGACTTCTTAGCAGSAAAGCTGACTCAGTATGGGCCCAAGACAGGCMC	487			
OY	2188	tttgtccatgtagggctcatctcagcagatgggggaagatttgacagatnacatgataybag	2247			
Db	488	TTTTGTCATGAGATGGGCTCACCTCCGATGGGAGTGTTTAATGATATACAAACAGCAG	547			
OY	2248	aaattcctaattccaatgygaagaalataaagcaqtaagatyltcaagcaqatatlactglt	2307			
Db	548	AAGTCTACTTATTCCAAAGSAAACCCCAACAGTAGAGSTTTACGACGCAATTACGGST	607			
OY	2308	acaaatgtagtaaaagatgltcaagygagcaagctgttacccaa---aagatgcacatlc	2364			
Db	608	AAAAATCAAGTTTCCGCGTGCCAGGAGGACAGTGTATCATTACGMAAGTGTGTATTC	667			
OY	2365	aataaagtaaaagagcttatagaanaaagatgtagtltgtlcccaaaccgccgcaagcg	2424			
Db	668	GACAAGATGACGGGCTGTATTAACACAAATTTGATTATGACAGATCCACCCAAAC	727			
OY	2425	gagaaggtctctatabatgtttgcacaacaatgttgatctctatagttgaattctgtcacgaa	2484			
Db	728	GAGAAGSGTTCATCATGTGTTTAACCAAAATATCATTTGTGTGGTTGATTTCTGRACAA	787			
OY	2485	caaaaaccaacaagaagctcccacaagaanaaatacaaaatlgcaatlcocgaagcaca	2544			
Db	788	AAAAATCACAATCAAGAAGACCCCAAAATGACCAAAACCAACGATGCATCTCGAAGCAG	847			
OY	2545	tygggaagtagtccgtgtagtcttgaggagctttaagaanaaccaactcetaatgaacaacagcca	2604			
Db	848	TGGGAAGTCATCCAGGAATCTGAGAGACTTCAAGCAAAACCACTCCATACAGCCACGCA	907			
OY	2605	ccaatctccacctctcaatgtctgcagaatlygaacaagaatltgtgtgtttagtocttiac	2664			
Db	908	CCTGCACCCACCTTCTCACTGCTGCAAAATTGGACAAAGAAATTGTGTGCTTAGTTCCTGAT	967			
OY	2665	aaatctygaagaatlygcagctcygtaaaccgctccaatgcactygaatlcgaacagcgcaagt	2724			
Db	968	AAGTCCGGGACATCGTCAACGATGATGCTGTGACCGAATGAAATACAGGCAAGCCGGCTT	1027			
OY	2725	ttctcctgtcagacaagttgagctgtgggtcctcgtggtltvggatlygtacatttgacaagtct	2784			
Db	1028	TTTCCTGCTGCGAGACTGTGGACAGGATCTCTGTGGGTGCGGATGGAGACTTGTGACAGTGCT	1087			
OY	2785	gcccatygaacaataatgaactatcacagataaaacgttgcgtgagtgacagygaaacactcgcc	2844			
Db	1088	GCTTATGTACCAAGGAACTCAACAGATTAAACAGTGGGTGCTGACAGAGATCTGCTGATC	1147			

OY	2845	aaaagattactccgaagaagctttaggaaggagagctccatctcgaaagggccttcgatactgcga	2904
Db	1148	AAGCACTTACCCACAGATATGTGCAGGAGGGACATCTATATGCTTGCGCTTCGGACACCA	1207
OY	2905	tttaactgatatgtygcacaacatttgccctgcttttccatgacacagaagttatbgyggagtg	2964
Db	1208	TTTTAC-----	1212
OY	2965	cgacagaagaatcccaaatctggccctctctgcccctgagcttagtgatctagaagaanaat	3024
Db	1213	-----AGTGATTAAGACAGATAT	1231
OY	3025	ccaactatgagatctgaaattgtgcgtcgcgcgaagataggaggaaagacaacataatagtygg	3084
Db	1232	CCAACCTATGGATCTGAAATTTGGCTGCTGACCGATGGGGAGACAAACCACTTAGCAGC	1291
OY	3085	tgctttaacagaggttcaacaagaatgtgtccatcatccacagctgcgttttgggcccct	3144
Db	1292	TGCTTTAGCTGGTGAAGCAGACGGGGCCGTCATTCATACAGTGGCCCTGGGACCGGCT	1351
OY	3145	gcagctcaagaactagagaagagctgtgtccaaatgacagagaag- tttaacagacatagcttc	3203
Db	1352	GGCCCTAAAGAGCTTGAGCAGCTGTGCTCAAAATGACAGAGGCGCTGACAGATATCTTTC	1411
OY	3204	agatcaagttccgaacaaatgagcctcatatgactcttttgggccccttcatcaagaatagtg	3263
Db	1412	GGATCAGGTTTCAGAACATATGGTCTTGTATGCTTTCGACAGACCTCTCCTCAAGGAAATGC	1471
OY	3264	agctgtctccagagctccacacacagctctgagtagaaggatataccctccagaagaagca	3323
Db	1472	GGCGATGGCTCAGCACCTCCATCCAGCTGGAGAGCAGGGAGTTAATTCACAAATAACCA	1531
OY	3324	gttgaatgaatgycacagtgatctgtygcacagcacocgttgggaaagagacacttgtttctat	3383
Db	1532	ATGAGATGAATGGCTCAAGTATCGTGGACAGCTGGGCGCAAGGAGACACTTGTTCCTAT	1591
OY	3384	caactgtgacaatgacagcctcccacaatccctctctgagatcccaagtgcagaagaag	3443
Db	1592	CACCTGGACACAGCATCTCTCTCAATATTTATCTTGGATTCACAGCGGAGTGGAAGAAA	1651
OY	3444	tgagcttctgtagtgcacaanaacccaataatgagcctacccccaatcccaagatgtgcia	3503
Db	1652	TGGTATTATCTAGACACCAACCACTAAGTGGCCTTACTCTCAAGTCCAGGACGGCTTAA	1711
OY	3504	ggttgcacttggaaatatacagctctgcgaagaagctacacaacacttgcagccttgcgttacc	3563
Db	1712	GGTGTGCTTTGGAAATACAGACATTCAGAGGAGCACTCAGACACTGTCACCTTGACTGCAC	1771
OY	3564	gtcccgtygcctcaatgctactccctgcctccaatatacagtgcacttccaaacgaacaaga	3623
Db	1772	CTCCCGTGACAGCAAGTCTCTCACTGGCTCTCTATTACAGACGCCGGTAAAGTGAATAAACAA	1831
OY	3624	caccagaagaattcccaagccctctgtgagttaatgaaatctgcgaagaagcctccc	3683
Db	1832	CACAGGAAATTTCCCGAGCCCTTAACAGTGTATGTCAACCACTTCGCAAGGAGCCTCGCC	1891
OY	3684	aattctcaagggcagtytccacaagcccttgattatcatcagtgaaatggaaaaaacyttactct	3743
Db	1892	TATTCTCAGGGCCAGCGCTCACACGCTTGATGTGAATGTGTGAATGGAATAAACAGTAACTCT	1951
OY	3744	ggaactacatgataatgagcaggtgcgtgactactaaagatgacgtygttactacaag	3803
Db	1952	GGAAATTAAGTGAATTAACGACAGGACGTCGCTCCGTCACCAAGATATGATGGTGTACTCAAG	2011
OY	3804	gtatttcaacatatagcacgaatggtgatgatacagtgfaaaagtyggggcctctggagg	3863
Db	2012	GTTTTTTACAGCTTTTATATCCAAATGGTAAATACACCGTTTAAATATATGGGCTCTGGAGG	2071
OY	3864	agtttaacgcagcagcagcagagtgatatacccccagcagatggagacacttataatccctgg	3923
Db	2072	AGTCACTTTCAGACAGACAGAGACGACGACACTCGCGAAGAACACAGCATATATATGATAGG	2131
OY	3924	cttgatctgagaatgataaataacatgtgaatccacaagaagccttgaatthaataagaatgta	3983

Db	2132	CTGGATTGGAGATGTGTAAGATTGAACCCACACCTCTCTGAAACTAGTT-----A	2185
QY	3984	tgctcacacaagaagtgctgtcttcagcagaacatcctcggaggctcatctgtgctc	4043
Db	2186	TGTTCAAGACACAGAGCTGTGCTTCAGCAGACAGATCTTACGGGGATCGTTGTGGCCAC	2245
QY	4044	tgatgtc---caaatgtctccatactgtatctctcccactggcccaatccaccgt	4100
Db	2246	CAATGTCCCGCAGACACCTCCATTCCTGTACCTCTTCCACCTGTCAATATCAGTACCT	2305
QY	4101	gaagcggaaatttcacggggcgagcttcaatctatctgtacttggacagctccctgggata	4160
Db	2306	GAAAGCCAGCATCCCAAGGGCAGAACCTGGTAATCTGACGTGTACGGCTCCGGGGGATGA	2365
QY	4161	ttatgacacatggacaacatcacaaaglatatcatcttgacaatgaatgaatcttgatct	4220
Db	2366	CTAGACACACGGAGAGAGCTTCCAACTACATCATCCAAATGACACACAGATATGTTGATCT	2425
QY	4221	cagagacaagttcaatgaatctctctcaagtgaaatactactgtctctcatcccaaggagc	4280
Db	2426	CAGGACACACTTCACACCTCACTCCCAAGGACATCCCGGCTTATCCCAAGAGGC	2485
QY	4281	caactctggaggagctcttcttctttaaaccagaacaatctacttcttgaataatggcacga	4340
Db	2486	CAGCTCGAGGAAATCTTCGAGTTTGAACGTGGGAGCAACGTTTTGGAAATGGCAGAGA	2545
QY	4341	tcttctcatgtcattcagctcgtgtbataagtcgactcgaatcagaatcagaatccaaat	4400
Db	2546	TATCTTATATGCTATCCAGAGCTGTGATGTAAGTCCAAATCTGAATTCAGAAATCTCCAAAT	2605
QY	4401	tgacaagatattcttgttattctctccacagagactccgcacagacacactagctcgtatga	4460
Db	2606	TGCAGCGGTGTGTGTGTATCATCCCGCTGAG-----GAGCGCGCATTCGCCGAGA	2656
QY	4461	aagctctgcctctgtctctataatattcatatcaaacagacacatctcctgcatcacattt	4520
Db	2657	CTCAACCTCCCTCTGTCTTCACTACATCAGCATCAACACACCATCTCTGGCATCCAGTGCT	2716
QY	4521	aaaattatgtggaagtgatgagagaacgtgcagctgtcataagcctag	4586
Db	2717	GAACTATATGTGAAATGCTAGGGGAATGCAAGTGAACATGAGTTTG	2765
RESULT	3		
AF101616			
LOCUS	AF101616	2703 bp	DNA
DEFINITION	AF101616	Human Homo sapiens genomic clone p1WB23.32SP6,	GSS
		sequence.	06-NOV-2000
ACCESSION	AF101616		
VERSION	AF101616.1	GI:4249719	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 2703)		
AUTHORS	Bepler G., O'Brian R.C., Kim Y.C., Schreiber G. and Pitterle D.M.		
TITLE	A 1.4-Mb high-resolution physical map and contig of chromosome		
	segment 11p15.5 and genes in the LOH1A metastasis suppressor		
	region		
JOURNAL	Genomes 55 (2), 164-175 (1999)		
MEDLINE	99134294		
COMMENT	Contact: Bepler G		
	Duke University Medical Center		
	Box 2610, MSRB, Room 117, Durham, NC 27710, USA		
	part of a 1.4 megabase contig including the LOH1A metastasis		
	suppressor region Bin B		
	Class: unknown.		
FEATURES	Location/Qualifiers		
source	1..2703		
	/organism="Homo sapiens"		

/db_xref="taxon:9606"
/map="11p15.5"
/clone="pWB233.32SP6"
/clone_lib="Human"
BASE COUNT 1057 a 573 c 483 g 590 t
ORIGIN

Query Match 17.7%; Score 808; DB 13; Length 2703;
Best Local Similarity 89.7%; Pred. No. 1.3e-166;
Matches 924; Conservative 0; Mismatches 50; Indels 56; Gaps 3;

QY 152 aattctaccagaggtacaaagaggaatctgtaccattccctctgaaagtattacaata 211
DB 1 AATTCTACCAAGAGGTACAAAGAGAGCTGTACCATCTCTGAAACTATTCACAA 60
QY 212 taagaaagaggaatccctcccttaactcgtttatagggccaatcatctgatataca 271
DB 61 TAGAAAAAGGGAATCCTCCCTAATCTTTATGAGGCCAGCATCATCTGTACCAA 120
QY 272 agcgggcaagagacacaaaccaaaagagaatttagacaatactttagaacaatg 331
DB 121 AGCCTCCAGAGACACAAACAAAAAGAGATGTTACACCAATATCCTGTGATGACATCG 180
QY 332 atgcaaaaatccctcaataaatactgcaaacccgaatccagcagcaataaagctta 391
DB 181 ATCCAAAAATCCCTCAATAAATACTGGCAACGTAATCCAGACACATCAAAAAAGCTTA 240
QY 392 tccacatgataagttgagttcatccctgagat----- 425
DB 241 TCCACCATGATCAAGTGGGCTTATCCTCCATGCAAGCTGTTCACATACGAAAT 300
QY 426 -----aaccaagacaaacacacatgatatctcaa 457
DB 301 CAATTAACATATCCAGCATATTAACAGAACCAAGACAAACACATATGATTATCTCAA 360
QY 458 tagatgcagaaagagcctttgacaaaattcaaacaccttcacgtctaaaacctcata 517
DB 361 TAGATCAGAAAAAGCCTTTGACAAAATTCACAAATGCTTCATGCTTAAATCTCAATA 420
QY 518 aattagatatgtaggacatactcaaaataaagaagctatctatgcaagccag 577
DB 421 AATTAGTATGATGGGACATATCTCAAAATTAAGAGCTATCTATGACAAACCCACAG 480
QY 578 ccaatatcatctgaatgtaggcaaaacttgaagcatccctcttgaanaacttgcacagac 637
DB 481 CCAATATCATACTGATGATGACAAAATAATGACATTCCTTTGAAAATGCGCAAGGC 540
QY 638 aggaatgcccctctcaacactctctatcaactagtcttggaaagtttgcagagga 697
DB 541 AGGATGCCCTCTCTCACCACCTCTATTCAACTAGTGTGTAAGTCTGGCCAGGCAA 600
QY 698 tttagcagagagaagaataaaggttttcaattagaaaaagagaagtcaattgtccc 757
DB 601 TCAGGAGAGAGAAAGAAATAAGGCAATTAAGAAAAAGAG-AGTAAATTGTGCC 659
QY 758 tgtttcaggtgacatgatgtgtatctacatgaataaacccattctctcagcccaatctcc 817
DB 660 TGTTCAGATGACATGATGTTATCTAGAAAACCCCATTTGTCGAGCCCAAAATCTCC 719
QY 818 ttaagttgataagcaacttcagaaagtctcagatatacaaatcaatgtcaaaaatcac 877
DB 720 TTAAAGCTGATTAAGCACTTCAGCAAAAGTCTCAGATATATAAATCAATGTCAAAAATTCAC 779
QY 878 aagcatctccatcaccaataacagagaagacccaatcatcgaatggaactccat 937
DB 780 AAGTATTTCTATCCACATTAACAGACAAACAGAGACCAATCATGAGTGAATCTCCAT 839
QY 938 tcaaatctgcttcaaaagagaataaataacacaggaatcccaacttaaggaagtgtgaag 997
DB 840 TCACATTTGCTTCAAAAGAGATAAATACTAGGAATCCAACTTACAAGGAGTGAAG 899
QY 998 acccttcaagagaactcaaacacactgctcaatgaataaagaagatacaacaat 1057

DB 900 ACCTCTCAAGAGAACTACAAACCACTGCTCATGAAATAAAGAGATACAAACAAAT 959
QY 1058 ggaagacattccatgctctcaatgtagaagaatacaatcatcgtgaanaatgccaactgc 1117
DB 960 GGAAGAACAATTCATGCTTTATGGGTAGGAAGATC-ATATGTAAGAAATGCTATCTGC 1018
QY 1118 ccaagattat 1127
DB 1019 CCAAGGTAAT 1028

RESULT 4
AUI22701 825 bp mRNA EST 23-OCT-2000
LOCUS AUI22701 MAMMAL Homo sapiens cDNA clone MAMMAL002934 5', mRNA
DEFINITION
sequence.
ACCESSION AUI22701
VERSION AUI22701.1 GI:10947417
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 825)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., and Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.
HRI human cDNA project
Unpublished (2000)
COMMENT
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3952
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES

Location/Qualifiers

source 1..825
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MAMMAL002934"
/clone_lib="MAMMAL"
/tissue_type="mammary gland"
/note="Vector: pME185FL3"
BASE COUNT 322 a 177 c 149 g 174 t 3 others
ORIGIN

Query Match 14.3%; Score 655; DB 10; Length 825;
Best Local Similarity 89.4%; Pred. No. 3.9e-133;
Matches 737; Conservative 0; Mismatches 33; Indels 54; Gaps 1;

QY 347 ataaatctgcaaacacgaatccagcagcacatacaaaagcttatccacatgatcaag 406
DB 1 ATTAATATCTGCGCAAAACCGAATCCAGCAGCATCAAAAAAGCTTATCCACCATGATCAAG 60
QY 407 tgggtctatccctggat----- 425
DB 61 TGGGCTTCATCCCTGGGATGCAAGGCTGTCATATACGAAATCAATTAATGTAATCC 120
QY 426 -----aaccaagaataaaacacacatgattatctcaatagatgcagaag 472
DB 121 AGCATATAACAGAGCCAAAGACAAAACACATGATTAATTCATGATGCAAGAAAAAG 180
QY 473 ccttgcacaaatcaaacaccttcatgctaaaacctcctaataatagatgatg 532
DB 181 CTTTGACAAAATTCACAAACCTTCATGCTTAATAAATCTCAATTAATTAATGATGATG 240

Oy	533	ggacatctcctaaataaataagagctctctctcctggcagaacgacagcctaatctactga	592
Db	241	GGAGCTATTTCAAAATATTAAGACTATCTATGCAAAACCACAGCAATATCTACTGA	300
Oy	593	atgggcaaaaacttgyaagcattccctcttgaaaacttggcacaagaacaggaatgcccctct	652
Db	301	ATGGGCAAAAACCTGGAAGCATCTCCCTTTGAAAACCTGGGACAGACAGGAGTCCCTCTCT	360
Oy	653	caccactcctcttcaacatagtttttggaaagttctctggccagggaacattagcagaagaag	712
Db	361	CACCGCTCTCACTCAACACTACTGTTGGAAAGTTCTGGCCAGGGCAATCAGGACGAGAGAAG	420
Oy	713	aaatcaagggttttccaattaggaagaagaagcaatattgctctcttttgcagttgca	772
Db	421	AAATTAAGGGTATTCAATTAGGAAGAAGGAGTCAAAATGTCCTCTGTTGCAGACGACA	480
Oy	773	tgattgtatacctagaaaaccccaattctctctcagcccaaatctccttaagctga	832
Db	481	TGATTGTTTATCTAGAAAACCCCATCTATCTCAGGCGCAAAATCTCTTAAGCTGATTAAGCA	540
Oy	833	acttcagcaaaagttctcaggaatacaaaatcaatgtacaaaatacaaaagcattcctataca	892
Db	541	ACTTCAGCAAAAGTCTCAGGATACAAAATCAATGTACAAAATTCACAAAGCATCTTATACA	600
Oy	893	ccaatcaacagagaacagagagagccaaatcatgtaattgtaactcccatcaaatgtctcaa	952
Db	601	CCAACACAGCAACAAACAGAGAGCCAAATCAATGGGTGAATCCCATTCACAAATTCCTTCAA	660
Oy	953	agagaataaatactcctaggaatcccaacttacaaggaagtgaagaagcctctcaagaga	1012
Db	661	AGAAATATTAATTACTTGAAGATCCAACTTACAAAGGATGTCAAGGACTCTTCAAGANGA	720
Oy	1013	actcaaaaccactgctcaatgaaataaanaagagaatacaaaatggaagaacattccat	1072
Db	721	ACTCAAAACCTCCTCCTCAGAGAAATTAAGAGANGACNCAACAAATGGAAGAATTCAT	780
Oy	1073	gctcatgtagtagaagaatcaatactgtgaaatagcgccactagt	1116
Db	781	GCTCATGGGTAGGAAGATCAATATCGGAAAGGCGCATCTG	824
RESULT	5		
LOCUS	BG622206	745 bp	18-APR-2001
DEFINITION	60264684.F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4768690 5', mRNA sequence.	EST	
ACCESSION	BG622206		
VERSION	BG622206.1	GI:13673577	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 745)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgaps-remail.nih.gov		
	Tissue Procurement: CLONTECH Laboratories, Inc.		
	cDNA Library Preparation: CLONTECH Laboratories, Inc.		
	DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	cDNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LDCM1632	row: m	column: 11
	High quality sequence stop: 734.		
FEATURES	location/Qualifiers		
source	1..745		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4768690"		

Query Match	Best Local Similarity	13.8%	Pred. 631:	DB 11:	Length 745:
Matches 661:	Conservative	0:	Mismatches 50:	Indels 0:	Gaps 0:
444	catgatatctcaatagatgacgaagaagccctttgacaacaattcaacaacccttcagt	503			
1	CATGATTTCTCAATAGATGCGAAAAGGCCCTTGACAAAATTCAACAGCCCTTCATGCT	60			
504	aaaaacccatlaaattagatatgtatggaacatactcaaaataataagaactatcta	563			
61	AAAACTCTCAATAGATTAGGTATGATGGATGATGATCTCAAAATPATTAAGACTATTTA	120			
564	tggcaaaagcccaagcaatatcatctgaatggtgcaaaacatcggaagcattccctttaa	623			
121	TGAAAACCCACAGCCCAATATCATATGTGATGAGGAAAAATCGAAGCATTTCCCTTTGAA	180			
624	aactgacaaagacaggaatgcctctctcaaccctccattcaacatagtttggagt	683			
181	AAGTGCACAAAGACAGGATGCCCTCTCCACCACTCCTATTCCACATAGTGTGGAAT	240			
664	tcctggccagggcaattaggcaggaagaanaaaagggtttccaattaggaaaaga	743			
241	TCTGCGCAGGCGCAATTAGGACAGGAAGAACTAAAGGGTATTCAATTAGAAAAAGGGA	300			
744	agtaaatgttcctcgtttgacagtgatagttgtatccctagaanaaacccattcttc	803			
301	AGTCAATTTGCCCTGTTTGACAGACACTGATTTGTATATTTAGAAAACCCCATCTCTC	360			
804	agcccaaaatctccttaagctatagaacaaacttcgaagaagctcgaataacaaatcaa	863			
361	AGCCCAAAATCTCCTTAAGCTGATTAAGCAACTTCACGAAAGTCTCAGGATACAAAATCAA	420			
864	tgtaaaaaaaltacaagaattcctatataccaataacagagaacagagagccaatcat	923			
421	TGTGCAAAAATCACAAAGCATTTCTTATACACCAATATATAGCAAAAGGAGGCAAAATCAT	480			
924	gaatgaactcccatcacaattgtcttcaagaagaataaatactcaggaaatccaacttc	983			
481	GAGTAACTCCCATTCACAAATTTGATTCAAAAGGAATTAATTAATCTTAAGAACTTTC	540			
984	aagggaatggaagagcccttcaagaagaacatacaaaacacgctccaatggaataaaga	1043			
541	AAAGGACTGGAAGGCTCTTTCAGGAGAACTTCAAAACCCTGCTCAATGAATGAAAAG	600			
1044	ggaatacaaaatggaagaacattccatgctcatgagtaggaagaatcaatatcgttaa	1103			
601	GGACACAAACAAATGGAAGACATTCATGCTCATGAGATGAGGAAGATCAATATCATGAA	660			
1104	aatggccatctgcgccaagatatgctgtagatataaaggatcatcaatagg	1154			
661	AATGGCATACTGCCCAAGTAAATTTACAGATTCAATGGCATGCCATCAAG	711			

BASE COUNT 285 a 159 c 133 g 168 t

ORIGIN

Query Match 13.8%; Pred. 631: DB 11: Length 745;

Best Local Similarity 93.0%;

Matches 661: Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Library: /lab_host="NIH_MGC_79" /lab_host="DH10B (T1 phage-resistant)" /note="Organ: Placentia; Vector: pNDR-LIB (Clontech); Site_1: SfilI (ggcgccctggcgc); Site_2: SfilI (ggcctatagcgc); Site_3: SfilI (ggcctatagcgc); Site_4: SfilI (ggcctatagcgc); Site_5: SfilI (ggcctatagcgc); Site_6: SfilI (ggcctatagcgc); Site_7: SfilI (ggcctatagcgc); Site_8: SfilI (ggcctatagcgc); Site_9: SfilI (ggcctatagcgc); Site_10: SfilI (ggcctatagcgc); Site_11: SfilI (ggcctatagcgc); Site_12: SfilI (ggcctatagcgc); Site_13: SfilI (ggcctatagcgc); Site_14: SfilI (ggcctatagcgc); Site_15: SfilI (ggcctatagcgc); Site_16: SfilI (ggcctatagcgc); Site_17: SfilI (ggcctatagcgc); Site_18: SfilI (ggcctatagcgc); Site_19: SfilI (ggcctatagcgc); Site_20: SfilI (ggcctatagcgc); Site_21: SfilI (ggcctatagcgc); Site_22: SfilI (ggcctatagcgc); Site_23: SfilI (ggcctatagcgc); Site_24: SfilI (ggcctatagcgc); Site_25: SfilI (ggcctatagcgc); Site_26: SfilI (ggcctatagcgc); Site_27: SfilI (ggcctatagcgc); Site_28: SfilI (ggcctatagcgc); Site_29: SfilI (ggcctatagcgc); Site_30: SfilI (ggcctatagcgc); Site_31: SfilI (ggcctatagcgc); Site_32: SfilI (ggcctatagcgc); Site_33: SfilI (ggcctatagcgc); Site_34: SfilI (ggcctatagcgc); Site_35: SfilI (ggcctatagcgc); Site_36: SfilI (ggcctatagcgc); Site_37: SfilI (ggcctatagcgc); Site_38: SfilI (ggcctatagcgc); Site_39: SfilI (ggcctatagcgc); Site_40: SfilI (ggcctatagcgc); Site_41: SfilI (ggcctatagcgc); Site_42: SfilI (ggcctatagcgc); Site_43: SfilI (ggcctatagcgc); Site_44: SfilI (ggcctatagcgc); Site_45: SfilI (ggcctatagcgc); Site_46: SfilI (ggcctatagcgc); Site_47: SfilI (ggcctatagcgc); Site_48: SfilI (ggcctatagcgc); Site_49: SfilI (ggcctatagcgc); Site_50: SfilI (ggcctatagcgc); Site_51: SfilI (ggcctatagcgc); Site_52: SfilI (ggcctatagcgc); Site_53: SfilI (ggcctatagcgc); Site_54: SfilI (ggcctatagcgc); Site_55: SfilI (ggcctatagcgc); Site_56: SfilI (ggcctatagcgc); Site_57: SfilI (ggcctatagcgc); Site_58: SfilI (ggcctatagcgc); Site_59: SfilI (ggcctatagcgc); Site_60: SfilI (ggcctatagcgc); Site_61: SfilI (ggcctatagcgc); Site_62: SfilI (ggcctatagcgc); Site_63: SfilI (ggcctatagcgc); Site_64: SfilI (ggcctatagcgc); Site_65: SfilI (ggcctatagcgc); Site_66: SfilI (ggcctatagcgc); Site_67: SfilI (ggcctatagcgc); Site_68: SfilI (ggcctatagcgc); Site_69: SfilI (ggcctatagcgc); Site_70: SfilI (ggcctatagcgc); Site_71: SfilI (ggcctatagcgc); Site_72: SfilI (ggcctatagcgc); Site_73: SfilI (ggcctatagcgc); Site_74: SfilI (ggcctatagcgc); Site_75: SfilI (ggcctatagcgc); Site_76: SfilI (ggcctatagcgc); Site_77: SfilI (ggcctatagcgc); Site_78: SfilI (ggcctatagcgc); Site_79: SfilI (ggcctatagcgc); Site_80: SfilI (ggcctatagcgc); Site_81: SfilI (ggcctatagcgc); Site_82: SfilI (ggcctatagcgc); Site_83: SfilI (ggcctatagcgc); Site_84: SfilI (ggcctatagcgc); Site_85: SfilI (ggcctatagcgc); Site_86: SfilI (ggcctatagcgc); Site_87: SfilI (ggcctatagcgc); Site_88: SfilI (ggcctatagcgc); Site_89: SfilI (ggcctatagcgc); Site_90: SfilI (ggcctatagcgc); Site_91: SfilI (ggcctatagcgc); Site_92: SfilI (ggcctatagcgc); Site_93: SfilI (ggcctatagcgc); Site_94: SfilI (ggcctatagcgc); Site_95: SfilI (ggcctatagcgc); Site_96: SfilI (ggcctatagcgc); Site_97: SfilI (ggcctatagcgc); Site_98: SfilI (ggcctatagcgc); Site_99: SfilI (ggcctatagcgc); Site_100: SfilI (ggcctatagcgc); Site_101: SfilI (ggcctatagcgc); Site_102: SfilI (ggcctatagcgc); Site_103: SfilI (ggcctatagcgc); Site_104: SfilI (ggcctatagcgc); Site_105: SfilI (ggcctatagcgc); Site_106: SfilI (ggcctatagcgc); Site_107: SfilI (ggcctatagcgc); Site_108: SfilI (ggcctatagcgc); Site_109: SfilI (ggcctatagcgc); Site_110: SfilI (ggcctatagcgc); Site_111: SfilI (ggcctatagcgc); Site_112: SfilI (ggcctatagcgc); Site_113: SfilI (ggcctatagcgc); Site_114: SfilI (ggcctatagcgc); Site_115: SfilI (ggcctatagcgc); Site_116: SfilI (ggcctatagcgc); Site_117: SfilI (ggcctatagcgc); Site_118: SfilI (ggcctatagcgc); Site_119: SfilI (ggcctatagcgc); Site_120: SfilI (ggcctatagcgc); Site_121: SfilI (ggcctatagcgc); Site_122: SfilI (ggcctatagcgc); Site_123: SfilI (ggcctatagcgc); Site_124: SfilI (ggcctatagcgc); Site_125: SfilI (ggcctatagcgc); Site_126: SfilI (ggcctatagcgc); Site_127: SfilI (ggcctatagcgc); Site_128: SfilI (

DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE	TITLE	JOURNAL	MEDLINE	COMMENT
HS_5574.A2.F10.T7A.RPCT-11 Human Male BAC Library Homo sapiens genomic clone Plate=1150 Col=20 Row=K, DNA sequence.	AF0748956	AF0748956	GI:5536114	human.	human.	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)	99380589		
1 (bases 1 to 943)				Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,J.									
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)									
High Throughput Sequencing Center									
University of Washington									
401 Queen Anne Avenue North, Seattle, WA 98109, USA									
Tel: (206) 616-3618									
Fax: (206) 616-3887									
Email: jwallace@u.washington.edu									
Clones are derived from the human BAC library RPCT-11. For BAC library availability, please contact Pieter de Jong									
(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (http://www.resgen.com). BAC end web Server: http://www.htsc.washington.edu									
Plate: 1150 row: K column: 20									
Seq primer: T7									
Class: BAC ends									
High quality sequence stop: 943.									
Location/Qualifiers									
1..943									
/organism="Homo sapiens"									
/db_xref="taxon:9606"									
/clone_plate=1150 Col=20 Row=K"									
/clone_lib="RPCT-11 Human Male BAC Library"									
/sex="male"									
/note="Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"									
BASE COUNT	345 a	220 c	159 g	215 t	4 others				
ORIGIN									
Query Match	13.8%	Score 630.4;	DB 13;	Length 943;					
Best Local Similarity	88.2%;	Pred. No. 9.8e-128;							
Matches 763; Conservative	0;	Mismatches 44;	Indels 58;	Gaps 5;					
20	catatacctctcccaagactaaaccggaagaagtgtgatactctgatatgacataacag	79							
23	CATACACCCCTCCCAAGACTTAACACGGAAGAGTTGAACTCTGTGAATGACTTAATAACAG	82							
80	gctcgatattgtgcaataatcaagagcttaccacaacaaagaagtcaggaacagatg	139							
83	GCTCGATATTGTGCAATAATACATTAAGTTACACCAACAAAAAAGTCCAGACCAAGT	142							
140	gattcacagctgaattctaccagaggtacaaaggaaggaactggtacattccctgtgaag	199							
143	GATTCACAGCCGAATTTCACGAGAGTGAAGAGAGAGCTGTACCAATTCCTTGGAAC	202							
200	tattacaatcaatagaagaagaaggaatccctccctaactcglttattatgaagccaatca	259							
203	TATTCCAAATCTATGAAAAAGAGGAAATCCTCCCTACCTCAATTTATAGGACCAATCA	262							
260	tcctgataccaaagccggcagagacacaaacaaagaagaattttgaccaatattctt	319							
263	TTCTGATATCCAAAGCCTGGCAGAGACCAACAAAAAAGAGATTTTGAACCAATATCC	322							

QY	320	tgatgaacattgtagcaaaaatccctcaataaatactctgcaaacggaatccagcgacga	379
Db	323	tgatgaacattgtagcaaaaatccctcaataaatactctgcaaacggaatccagcgacga	382
QY	380	lcaaaaagcttataccacatgataagctgagcttccatccctggagat-----	425
Db	383	tcAAAAGCTTATCCACCATGATGATCAAGTGCGCTTATCCCTGGATGCAAGCGCTGTTCA	442
QY	426	-----gaccaaagacaaaaccaca	445
Db	443	ACATACACAAATCAATTAATGTATCCAGCATATTAACAGAACCAAAACACAAAACACA	502
QY	446	tgattatctcaatagatgcaagaagaagcccttgacaaaatccaaca--cccttcatgcta	504
Db	503	tgattatctcaatagatgcaagaagaagcccttgacaaaatccaacaagcccttcatgcta	562
QY	505	aaaaccctcaataaataagatattgattgtaggacatactccaataataagaagctcat	564
Db	563	AAAATCTTCATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	622
QY	565	ggcaaaagccacagccaatatatactatgataatgggcaaaaactgggaagcccttcccttgaa	624
Db	623	GACCAACCCACACGCCAATATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	682
QY	625	actggcacaaagcagagatgcccctctctccacacccctcatccaatagattttg--gaagt	683
Db	683	ACTGGCACAAAGCAGAGATGCCCTCTCTCCACACGCCCTCATTTCAATATGATGATGAT	742
QY	684	tctggccagggcaattagagcagagagaagaataaagggtttccaatagagaagaaga	743
Db	743	TCTGGCCAGNGCAATTAAGGACAGAAAGAAATTAAGGCAATTAATTAAGGAAAGAGA	802
QY	744	agtcgaattgtccctgtttgcaggtggaatgattgataccctagaagaaccatctctc	803
Db	803	AGTCGAATTTGCCGTGTTGCAGAT-ACATAATTTGATATTGAGAAATCCCATCTCTC	861
QY	804	agcccaaaatctccttaagctgata	828
Db	862	AGCCC-AAATCTCCTTAAGCTGATA	885
RESULT	7		
LOCUS	A0629742	810 bp	DNA
DEFINITION	RPci-11-467B15.TV	RPci-11	Homo sapiens genomic clone RPci-11-467B15
ACCESSION	A0629742		
VERSION	A0629742.1	GI:5092377	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	Zhao, S., Adams, M. D., Nierman, W., Malek, J., de Jong, P. and Venter, J. C.		
JOURNAL	Use of BAC End Sequences from Library RPci-11 for Sequence-Ready		
COMMENT	Map Building Unpublished (1997) Other GSS: RPci-11-467B15.TJ Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbe@tigr.org Clones are derived from the human BAC library RPci-11. For BAC library availability, please contact Pieter de Jong (pieter@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (inforesgen.com). BAC end search page:		

Db	259	TTCAACATAGTGTGTGGAGTCTTGCCCGAGGGCATTAGCGAGGAGAAATAAAGGT	358
QY	724	tttcaattaggaaaaaggaagtcgaattgtccctgtttcgaggtgacatgattgtatc	783
Db	359	ATTCAATTAGGAAAAAGGAGATGCAAAATGTGCCGTGTTGACAGACATCATTTGTAT	418
QY	784	ctagaataaccccatctcttcagcccaaatctccttlaagtcgataagcaacttcaacaa	843
Db	419	CTAGAAAAACCCATTGTCTTCAGCCCAAAATGCTTAAAGCTGATAGCACTTTCAGCAG	478
QY	844	gtctcagatatcaaatcaatgttcaaaaaatcaagaattccatcaccaataacaga	903
Db	479	GTCCTAGGATTCAAAATCAATGTACAAAATACACAGCATTTATACCAATTAACAGA	538
QY	904	gaacacagagaccgaatcatgaaatgaactcccatcttcaacttgccttcaagaagataaa	963
Db	539	CAACACAGAGACCCAAATATATGATGATCACTCCCTTCACCATTTGCTTCAAGAGATTAAA	598
QY	964	tacttaagatccaacttacaagggatgtgaaggaactcttcaaggaagaaactacaaca	1023
Db	599	TACCTAGAAATGCCAATTACAAAGGATGTGAAGGACCTTNCAGAGAACTTACAAACA	658
QY	1024	ctgtctcaatgaataaaagagatatacaaaaatlygaagaaacttccatgtctcatgata	1083
Db	659	CTGCTCATGTGAATTAAGAGAGATACCAACAATTGGAAGAACATTCCTCATTGGGTA	718
QY	1084	ggaagaatccaatatcgtyaaaaatgycgcatactgcccaagattatgcta	1131
Db	719	GGAAACAATCANTATTGTGAAAAATGCC-CTACTGCCCAAGTTATTACAA	765

RESULT	9
LOCUS	AU137259
DEFINITION	AU137259 893 bp mRNA EST 24-OCT-2000 AU137259 PLACE1 Homo sapiens CDNA clone PLACE106114 5', mRNA sequence.
ACCESSION	AU137259
VERSION	AU137259
KEYWORDS	AU137259.1 GI:10997798 EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 893) Oie,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yanamoto,J., Wakamatsu,A., Nakamura,Y., Negai,T., Sugano,S. and Isogai,T.
TITLE	HRI human cDNA project
JOURNAL	Unpublished (2000)
COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel.: 81-438-52-3951 Fax: 81-438-52-3952 Email: genomcs@hri.co.jp HRI human CDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute. Location/Qualifiers 1..893
FEATURES	
source	

BASE COUNT	ORIGIN	337 a	187 c	167 g	195 t	7 others
/organism="Homo sapiens"						
/db_xref="taxon:9606"						
/clone="PLACE1006114"						
/clone_1b="PLACE1"						
/tissue_type="placenta"						
/note="Vector:PM18SFL"						

Query Match	13.5%;	Score 617.8;	DB 10;	Length 893;
Best Local Similarity	87.1%;	Pred. No. 5.6e-125;		
Matches 765;	Conservative 0;	Mismatches 54;	Indels 59;	Gaps 6;

OY	1	atgataaattcctcaacacatacaccctccccaagactaacacaggaagaagttagct	60
Db	1.7	ATGGATACATATCTCTGCACACTACACTCTCCCAAGACTTAACCAAGAAAGTTGTAATCT	76
OY	61	ctgaatgaccataaacaagctctgatatgttggaataaacaagctcttaccacaaca	120
Db	61	ctgaatgaccataaacaagctctgatatgttggaataaacaagctcttaccacaaca	120
Db	77	CTGATATACCAACCAATPAACAGGAGCTGGAATAATGTGGCAATTAATCACTATTTCACACCAAA	136
OY	121	aagagtcacagaccagatgtgatacaagctgaaattctacacagagtgatacagaggaactc	180
Db	137	AAGAGTCCAGGACCAAGATGGAATTACAGCTGTAATTCTACAGAGGTCAAGAGGAAGACTG	196
OY	181	gtaccattccctctgaaagatatacaatcaatagaaaaagaggcaatccctccactcg	240
Db	197	GTACCATTCCTCTTGAAACATATTCATCAATATAGAAAAAGAGGAATCTCTCCATCTCA	256
OY	241	tttatgagggccaatcatalccttgatataccaaagccgggcaagagacaaacccaataaag	300
Db	257	TTTTATGAGGGCACATCATCTTATACCAAGCGGGGCGAGACACAAACCAAAAAAGAG	316
OY	301	aatttgaaccaatctcttgatagaacatttgatgcaaaaatccccaataaatactgca	360
Db	317	AATTTTAGACCAATATCTCTTGATATACATTAATGATGCAAAAAATCCTCAATATAATATCTGCA	376
OY	361	aaccgaatccagcagcacatcaaaaagcttatccaccaatgataagtgggcttccact	420
Db	377	AACGGAATCCAGCAGCAGCATCAAAAGCTTATCCACCATGTCMAAGGGCTTCATCCCT	436
OY	421	ggag-----a 426	
Db	437	GGGATGCAAGGCTGGTTCATATATACGAGTCATTAATGTAAATCCAGCATATATAACGA	496
OY	427	acgaagaacaaaacacatgatatcttcaatagaatgcagaaaagccttgacaant	486
Db	497	GCCAAAGCAAAAACCAACATGATTTATCTCATATATATCAGAAAAAGCCTTTGACAAAATT	556
OY	487	caacaacccctcactgctaaaaaacccctaataaactagatataatgtgagacatactca	546
Db	557	CAACACCCCTTCATGCTATAAAACCTCAATTAATTAAGTATTGATGGGATGTATTTCAAA	616
OY	547	ataataagactatctatgcaaaagccacagccaatatcatactgaaatggygcaaaaact	605
Db	617	ATAATAGAGGTATCTATGACAAAACCAACGACATATCAATGATGAGTGGGCAAAAACG	676
OY	606	ggaagccttcccttgaaaaactgcaaaaaga-ggaatgcctctcttaccactctat	664
Db	677	GGAAGCACTCCCTTGAAAMCTGGACACAGCAAGGAGATGCCCTCTCTCAACGCTCTAT	736
OY	665	tcaacatagcttggaagtlctgcgcaggggaactgaagc-aggagaagaaataaagyt	723
Db	737	TCACATATGTGTGGAGTTCTGGGCAAGGAATACAGGCAAGGAAGGAATAAAGGT	796
OY	724	ttcaataagaaaaaggaagtcacaatgtgccctgtgtgcaggtgacat-gattgtata	782
Db	797	ATTCAATTTGGGAAAAAGAGGAAGTC-AATTGGCCCTGTGTTGCAGACGACATGATTTGGTTA	855
OY	783	ctctgaanaaccacatctctcagcccaaatctccta 820	
Db	856	NCTAGAAAACTCATATNTAAGGCCAAAAATNNCCCTTA 893	

RESULT	10		
LOCUS	AQ752204		
DEFINITION	<p>AQ752204 863 bp DNA GSS 19-JUL-1999 HS-5571_B2_H12-T7A_RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1147 Col=24 Row=P, DNA sequence.</p>		
ACCESSION	AQ752204		
VERSION	AQ752204.1	GI:5539362	
KEYWORDS	GSS.		

QY	457	atagctgcagaagaagcctttgacaaatltcaacaaccttcattatgctcttaaaaccttcaat	516
Db	374	ATAGATGAGAGAAAAAGCCCTTTGGACAAATTTCAACAAACCCCTTCATCTCAAAACTCTCAT	433
QY	517	aattagatattgtagggacatctcaaatataaagaagcatctatgagcaaacgaca	576
Db	434	AAATTAGTATTGATGGGNCATATTTCAAATAATATAGAGGCTATCTATGACAAACCCACA	493
QY	577	gccaatcatcactgaatgagcaaaaacttgaagaattcccttltgaaacttgcacaga	636
Db	494	GCCATATCATCTACTTAACAGGGGCAAAAACCTGGAGAGCATTCCTTTGAAAACTGGACACAGA	553
QY	637	caggatgacctctcttcaccacctctatccaacatagtttggaaattctggccagga	696
Db	554	CAGGATGACCCCTCTCTCCACACTCTCTATTTCAACATATGTTGGAGTTCTGGCCAGGGCA	613
QY	697	attagcgaggaagaagaataaagggtttcaattaggaagaagaggaatgaattgtcc	756
Db	614	ATTAGCGAGGAAGGGAATTA--GGGATTCATTAGGGAAGAGAACTCAAAATTGTC	671
QY	757	ctgtttgcaagtgatgatgtatgtatatacttagaanaaccctatctctcagcccaaatctc	816
Db	672	CTGTTTGCAGAGACATGATGTATATCTAGAANACCCACATGCTGCAGCCCAAAATCTT	731
QY	817	cttaagctgataagcaacttcagcaaaagtctcaggatatacaaatcaatgatacaaatca	876
Db	732	CTTAAGCTGATATAGCAACCTTAGCAAGCTCTCAGGATCCAAATCAATGTACAAAAATTA	791
QY	877	caagattccctctacaccatatacagaagaagaagcaaatcatgaataactccca	936
Db	792	CAAGCTTTTATTTCNCCACACCCAGCACAGACAGAGCCCAATTCGTGAGTGACTTCCA	851
QY	937	ttcacacattgct 948	
Db	852	TTTCACATGTGT 863	
RESULT 11			
LOCUS	AM361534/c		
DEFINITION	QV2-CT0261-261099-011-f01 CT0261 Homo sapiens cDNA, mRNA sequence.		
ACCESSION	AM361534		
VERSION	AM361534.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 633)		
JOURNAL	HCBP http://www.ludwig.org.br/ORESTES .		
COMMENT	The FAPESP/LICR Human Cancer Genome Project Unpublished (1999) Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl-QV2&ct-QV2-CT0261-261099-011-f01&ts=1999-10-26&td=1) Seq primer: puc 18 forward High quality sequence start: 37 High quality sequence stop: 632. Location/Qualifiers 1..633 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="CT0261" /dev_stage="Adult"		
FEATURES			
source			

/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /16 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 140 a 156 c 166 g 191 t
ORIGIN

Query Match 13.4%; Score 611; DB 10; Length 653;
Best Local Similarity 98.9%; Pred. No. 1.7e-123;
Matches 656; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY 3317 acagcagatgagatgacagatgacagatgacagacacgctgggaagagacattgtc 3376
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 644 ACAGCCAGTGA -GAATGCACAGTATCTGACACACCGTGGAAAGACACTTGT 586
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3377 tctatcaccttgacaatgacagcctcccaatccttctctggatccagtgacaga 3436
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 585 TTCTTGACCTGGACAAACGACGCTCCCAATCTTCTGGATCCAGTGACAGA 526
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3437 agaaaggtgcttctgtgagtgacaaaacacaaatgctctcctcaatccagga 3496
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 525 AGCAAGGTGCTTTGATGACAAAAACAAAGGCTACCTCAATCCAGCA 466
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3497 ttgctaagctgacacttgaaatacagctcgaagacagacacacacacacacac 3556
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 465 TTCTTAAGTGGACACTTGAAATACAGTCTGCACAGCACTCAACCTTGACCTGA 406
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3557 ctgtcacctccgtgctgccaatgctacccctcccaatgacatgacttccaaaga 3616
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 405 CTGTCAAGTCCGCTGCTCAATGCTACCTGCTCAATTAAGTGAATCCAAAGCA 346
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3617 acagaagacacgaatcccaagcctctgtgagtttatgcaaatatccgcaagag 3676
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 345 ACAGAGACACCAAAATTCACAGCCCTGTGATGCAAAATTCGCAAGAG 286
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3677 ccccccacatcccaagcagctgacagcctcgtatgacatgacatgacatgacat 3736
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 285 CCGCCCAATTCACAGGCGACAGTGCACACCTGATTGAATTAAGTGAAGCA 226
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3737 ttaacctggaactcgtgataatgagcaggtgctgactgactaagatgacgtgtct 3796
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 225 TTACCTTGGAACTACTGGAATGAGACAGTGTCTGATCTACTAAGATGACGAGTCT 166
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3797 actcaagatattcacactatgacagatgtagatgacatgtaaaagtgcggtctc 3856
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 165 ACTCAAGATTTTCACAACTTATGACACGAATGTAGATACAGTGTAAAGTGCAGGCTC 106
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3857 tgggaagagttaacgacagcagcagagatgatacccccagagtgagacacgtgaca 3916
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 105 TGGGAGAGATTACGACGACGACGAGAGTGTATACCCCGACGAGTGGAGCACTGTACA 46
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3917 taactgg-ctgagatgagaatgataaatacaatgaaatccac 3958
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 45 TACCTGACTGATGGAATGATGAATATCAATGCAATCTCTC 3

RESULT 12
AO896432/c 872 bp DNA GSS 10-NOV-1999
LOCUS HS.3134.A1.A11-T7C CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate-3134 Col=21 Row=4, DNA sequence.
ACCESSION AO896432
VERSION AO896432.1 GI:6352622
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 872)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3134 row: A column: 21
Seq primer: 77
Class: BAC ends
High quality sequence stop: 872.
Location/Qualifiers
1. 872

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-3134 Col=21 Row=A"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 197 a 155 c 172 g 342 t 6 others
ORIGIN

Query Match 13.4%; Score 610; DB 13; Length 872;
Best Local Similarity 93.5%; Pred. No. 2.9e-123;
Matches 657; Conservative 0; Mismatches 43; Indels 3; Gaps 2;

QY 424 ataacaagaacaaaacacatgatctcaatagatgcagaaaagccttgacaa 483
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 710 AGAACCAAGACAAACCAATGATATCTCAATGATGAGAAAGAGCCTTGGACAA 651
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 484 attcaacaaccccttcacgtcctcaaaaacccctcaataatgatatgttgagacatctc 543
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 650 ATCCACCAACGGCTTCATGCTTCAAAAACCTCAATTAATTAAGTATGAGGACGTATGTC 591
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 544 aaataataagagctatctatggcaagccacagcccaatctcatctgaaatgggcaaaa 603
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 590 CAATATTAAGACTATCTATGACAAACCCACAGCCCAATTCATGATGATGGCAANMA 531
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 604 ctggaagcattcccttgaaaactgycacagaacaggaatgcccctctcactcactccta 663
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 530 CTGGAAGAAATTCCTTTGAANAACCTGGACAAAGACAGGATGCCCTCTCACCACCTCTTA 471
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 664 ttaacaatagattttgagaatctggtccaggtcaattaggcaggaagaagaataaaggt 723
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 470 TTCAACATAGTGTGGAAAGTGTGGCCAGGCAATACAGCAGAGAGAAAGAAATMAAGGT 411
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 724 ttccaattagaagaagaagatcaaatgttcctgtgttgacgtgacatgattgtatc 783
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 410 ATTCAATTAGAAAAGAGAGAGATCANATTGTC -GTTTCAATGACATGATTTATATAT 353
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 784 ctagaanaaccccaattcctcagcccaaatctcctaagctgataagcaacttcagcaaa 843
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 352 CTGAAGAAACCCCAATTTGCTCAGGCCCAAAATCTCCTTACGCTGATTAAGCAACTTCAGCAAA 293
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 844 gtctcagatacaaaaatcaatgtacaaaataacgaagcatctctataaccca-ataacg 902
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 292 GTCTCAGGATACAAAATCAATGTGCAAAAATCAACAGCATCTTATATACACCAATATATAA 263
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 903 agaaacagagcccaatcagatgataaactccatccatcaaatgcttcaagaagaataaa 962
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 232 ACAAAAAGAGCCCAATCTATGAGTACTCCATTTCACATTTGCTTCAAAAGGAATATA 173

QY 963 ataccaggaaatccaacttacaaggagatgagagacccttccaaggaatacacaacc 1022
 |||||||
 DB 172 ATACTAGAAATCAACTTACAGGGAATGAGGAGCTCTTCAAGGAGACTACAAAC 113
 |||||||
 QY 1023 acyctcaatgaataaagaaggatatacaacaatgagaacatctcatgctcatggt 1082
 |||||||
 DB 112 ACTGCTCAATGAATTAAGAGAGATGAGAAATGAAGAAATTCATGCTCATGGGT 53
 |||||||
 QY 1083 aggaagaatcaatctcgtggaatgagccatctgcccagaatt 1125
 |||||||
 DB 52 AGGAAGATCAATAGTGTGAACATGACATCTGCCCAAGGT 10
 |||||||

RESULT 13
 A0747583 860 bp DNA GSS 19-JUL-1999
 LOCUS HE_5353_A2.A06_77A RPCI-11 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate-1109 Col-12 Row-A, DNA sequence.
 ACCESSION A0747583
 VERSION A0747583.1 GI:5534741
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 860)
 AUTHORS Mahaitas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahaitas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Research Genetics (info@resgen.com). BAC end Web Server:
 http://www.htsc.washington.edu
 Plate: 1109 row: A column: 12
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 860.

FEATURES
 source location/Qualifiers
 1..860
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-1109 Col-12 Row-A"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methyase. Size selected DNA was cloned into the
 pBAC3.6 vector at EcoRI sites"

BASE COUNT 322 a 190 c 151 g 197 t
 ORIGIN

Query Match 13.3%; Score 608.6; DB 13; Length 860;
 Best Local Similarity 87.9%; Pred. No. 5.8e-123;
 Matches 743; Conservative 0; Mismatches 44; Indels 58; Gaps 5;

QY 153 ataccagaggtacaa-ggagaagactgatacctcctctgagaagtatacacaac 211
 |||||||
 DB 19 ATTCTACAGAGGTTCAAGGAGAACTGTACCTCTCTGAACTATTCACATCAA 78
 |||||||

QY 212 laggaaagagcaatccctccctactcgttttatgagggcaacatcctcgtatccaa 271
 |||||||
 DB 79 TAGAATAAGAGGAGTCTCTCCCTTACATCTTTTATGAGGCGACATCATCTGTATACAA 138
 |||||||
 QY 272 agccggcagagacacacacacaaagagaattttagacaaatctctgatacatg 331
 |||||||
 DB 139 AGCTGAGCAGAGACACACACAAAGAGAAATTTAGACCAATATCTGTGATGACATTTG 198
 |||||||
 QY 332 atgcacaaatccctcaataaataactgcaaacccgaatccagcagacatcaaaaagctta 391
 |||||||
 DB 199 ATGCACAAATCTCTCAATTAATTAATCTGCAACCCGATCCA-TTGTACATCAAAAAGCTTA 257
 |||||||
 QY 392 tccaccatgataagatgggtgttccatccctggat----- 425
 |||||||
 DB 258 TCCACCATGATCAATGAGGCTTCATCCCTGGGATGCAGGCTGTTCAATACATGCAT 317
 |||||||
 QY 426 -----aaccaagacaaaacacacatgatatcaca 457
 |||||||
 DB 318 CAATTAATGTATCCAGCATATTAACAGAGCCAAAGACAAACCAACATGATATCTCA 377
 |||||||
 QY 458 tagatgcagaaaggcccttgacaaaatccaaccctcatgcttaaaaccctcaata 517
 |||||||
 DB 378 TAGATGCAGAAAGGCCCTTGCACAAATTCACAA-CTTCAATGCTAAAAACTCAATA 436
 |||||||
 QY 518 aattgatattgattggacatatcccaaatataagagctctctatgcaagccacag 577
 |||||||
 DB 437 AATTGAGTATGATGGACATATTTCAAAATCATTAAGAGCTATGTATGACAAACCCACAG 496
 |||||||
 QY 578 ccaatatatactgataatggcacaagaactggaacatccctcttgaaaaactgcacagaag 637
 |||||||
 DB 497 CCAATATCATCATGTAATGGGCAAAAACGAGAAAGCATTCCTTTGAACCTGGCACAAGAC 556
 |||||||
 QY 638 aggaatgccctctctaccactcctataccaatagtttggaaagttctgacagggcaa 697
 |||||||
 DB 557 AGGAGTGGCCCTCTCTCACCGCTCCTATTCACATATGTTGGAAAGTCTGGCAGGCA 616
 |||||||
 QY 698 ttaggcagagagaagaataaagggtttcaattgagaaaagaggaagtcgaattgtccc 757
 |||||||
 DB 617 TTAGGCGAGAGAGAAATTAAGGGTATTCATTAAGAAAAGAGAGCAATTTGCTCC 676
 |||||||
 QY 758 tcttcagagtgatctgattgataccttagaagaacccaatctctcagcccaaatctcc 817
 |||||||
 DB 677 TGTTCGAGATGACATGATGTATATCTAGAAACCCCATTTCTGCA-CCCAAAATCTCC 735
 |||||||
 QY 818 ttaagctataagcaactcagcaagttcaggaatacaaaatcaatgatacaaaatcaac 877
 |||||||
 DB 736 TTAACTGATTAAGCAACTTACACCAAGTCTCAGATACAAATCAATGATCAAAAATCAC 795
 |||||||
 QY 878 aagctctctatacccaataacagagaacagagagccaaatctgataaacctccat 937
 |||||||
 DB 796 AGGCTTCTTTATACACCAACCAACCAAGAGAGCCAAATGAGTGCATCTTCAT 855
 |||||||
 QY 938 tcaca 942
 |||||||
 DB 856 TCACA 860
 |||||||

RESULT 14
 A0636664 796 bp DNA GSS 17-JUN-1999
 LOCUS A0636664
 DEFINITION RPCI-11-479A11.TV RPCI-11 Homo sapiens genomic clone RPCI-11-479A11
 , DNA sequence.
 ACCESSION A0636664
 VERSION A0636664
 KEYWORDS A0636664.1 GI:5099299
 GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 796)
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
 ,J.C.

FEATURES	Source	Location/Qualifiers
Journal Comment	Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building Unpublished (1997) Other_GSSs: RPCI-11-479A11.TV Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbeetlgr.org Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pietere@jng.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet. cs (info@resgen.com). BAC end search page: http://www.tlgr.org/tldb/hungen/bac_end_search/bac_end_search.html. Seq primer: T7 Class: BAC ends.	
BASE COUNT	302 a 178 c 131 g 185 t	
ORIGIN		
Query Match	13.3%; Score 606; DB 13; Length 796;	
Best Local Similarity	87.6%; Pred. No. 2.1e-122;	
Matches	697; Conservative 0; Mismatches 45; Indels 54; Gaps 1.	
Y	153 attctaccagaggtacaagaggaactcgttacccctcctctgaagatatacaatcaat	212
Db	1 ATTCTCCAGAGAGTACAAAGAGAGACTGGTACCATTCTCTGAAACATATCCATCAAT	60
Y	213 agaaaaagaggcaatccctccctaactcgtttatagaaggccaacatccttgatccaaa	272
Db	61 AGAAAAATAGACAATTCCTCCCTACATCTTATAGAGCTACATCATCTGATACCAAA	120
Y	273 gccggcgagagacacaacccaagaagagatttagaccatatacttgatgaacatga	332
Db	121 GCCTGGCAGAGACACAAACAAAAGAGAAATTTAGACCATAATCTCCGTGATGACATCA	180
Y	333 tgcacaaatcctcaataaatactcgcacaaaccgnaatccagacgacacatcaaaaagctat	392
Db	181 TGCAAAATVCCATCAATTAATAATCTGGCAAAACCAATCCAGCAACATATCAAAAAGCTTAT	240
Y	393 ccaccatgatcaatgaggccttcattccctggat	425
Db	241 CCACCATGATCAAGTGGGTTTCATCCCTGGGATGCGAAGGCTGGTTCACTATGCAAAATC	300
Y	426 -----aaaccaagaacaaaacacatgatatctcat	458
Db	301 AATAAAGCTATCCAGTGTATTAACAGANCCACAGACAAAACCAATGATATATCTCAAT	360
Y	459 agatgcagaaaaagcctcttgacaanaatcaaacaccccttcattcgtctaaaacctcaataa	518
Db	361 AGATGCACAAAAAGCCTTTGACAAAATTCACAAAGCCCTCCATGCTTAATAAACCTCTCATATA	420
Y	519 attgatatgatgagacatatctcaaaaataaagagctatctatgagaaagcagcagc	578
Db	421 ATTAGGTATTGTGACACATATCTCAAAAATAATAGAGCTATCTACGACAAACCAAGC	480
Y	579 caatatctactgaatggcacaanaactcggagcattccctttgaaanaactcggcacaagaca	638
Db	481 CAATATCTATCTGATGAGGCAAAAATCTGAAACCATCTCCCTTTGAAAACCTGGCACAAGACA	540

QY	639	ggagatgccctctctaccaccticctctttaaacaatgatttggaaatcttgcgcaagggaat	698
Db	541	GGGATGCCCTCTCTCTACCACTCTCTATTCACATAGCTGGAAAGTTCTGGCCAGGGAAT	600
QY	699	taggcagagagaagaataaaggcttctcaattaggaanaagggaagtcacaattgcct	758
Db	601	TAGGCAGCAGCAAGCAATTAAAGCGTATTCATATTAGGAANAAGATGAAGTCAATTGTCCCT	660
QY	759	gtttcgagtgacatgatgtatctaccctagaanaaccacatctctcctagcccaaatctct	818
Db	661	GTTTCAGATGACATGATGTATATCTAGAAAAGCCATCATCTCAGCCCAAAATCTCCT	720
QY	819	taagcttgataaacaacttcaagaagcttcagagatcacataaatacaatgataaanaactaca	878
Db	721	TAAAGCTGATTAACCACTTCAGCAAGCTCTCATGATATCAATAATGAATGTGCCAAACTACA	780
QY	879	agcattcctatacc	894
Db	781	AGCATTCCTATATACACC	796
RESULT	15		
LOCUS	AQ058259		
DEFINITION	CTT-HSP-2245F5.1R CTT-HSP Homo sapiens genomic clone 2345F5, DNA sequence.		
ACCESSION	AQ058259		
VERSION	AQ058259.1	GI:3354785	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 759) Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C. Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)		
TITLE	Unpublished (1998)		
JOURNAL	Other_GSSs: CTT-HSP-2345F5.TF		
COMMENT	Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html. Seq primer: M13 Reverse Class: BAC ends.		
FEATURES	Location/Qualifiers		
source	1..759		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="2345F5"		
	/clone_lib="CTT-HSP"		
	/sex="Male"		
	/cell_type="Sperm"		
	/note="Vector: pbeloBAC11; Site_1: HindIII; Site_2: HindIII"		
BASE COUNT	296 a 158 c 136 g 169 t		
ORIGIN			
Query Match	13.3%: Score 605.4; DB 13; Length 759;		
Best Local Similarity	95.6%: Pred. No. 2.9e-122;		
Matches 634; Conservative	0; Mismatches 26; Indels 3; Gaps 1;		
QY	424	ataaccaagaacaaacacatgatttctcaatagatgcgaaaaagcctttagacaa	483

```
Db 100 AGAACCAAGACAAAACACATGATATCTCAATAGATGAGAAAAGCCTTGACAAA 159
QY 484 attcaacacccctcatgcttaaaaaacctaaataatagatatgaggacatatc 543
Db 160 ATTCAACAGCCCTTCATGCTAAACTCTCAATAATTAGGATGATGGACGTATCTC 219
QY 544 aaaaataaagagctatctatgcaagcacaagccaalacatactgatatggcaaaa 603
Db 220 AAAATTAATTAAGAGCTATCTATGACAAACCAAGCATATCATTTGATGGCAAAA 279
QY 604 ctggaagcattcccttggaaaactgysaaaagaaaggaatgcccctctcaaccactccta 663
Db 280 CTGGAAGCATTCCTTTGAAAACGGGCACAAGACAGGGATGCCCTCTCAACACTCCTA 339
QY 664 ttcaacataagtttggaaagtctgcccaggccaattagcaggagaagaagaataaaggt 723
Db 340 TTCAACATATGTTGGATGTGCTGGCCAGGCAATTAGGCAAGAGAAATTAAGGGT 399
QY 724 ttcaataggaaaagaaagaaagtcgaattgtccctgttggcagtgacatgatatc 783
Db 400 ATTCAATTAGGAAAAGAGAGAGTCAATTGTCCGTGTTCAGATGACATGATTGTATAT 459
QY 784 ctgaaaaaaccatctctcagcccaaatctccttaagctgataagcaacttcagcaaa 843
Db 460 TTAGAAAAACCCATGCTCTCAGCCCAAAATCTC--AAGCTGATCAGCACTTCAGCAAA 516
QY 844 gtctcagagatacaaatcaatgtaacaaaatcacaagaatctccctatacacaataaaga 903
Db 517 GTCTCAGAGATACAAAATCAATGTAACAAAATCACAAGCATTCTTATACCAATTAACAGA 576
QY 904 gaacaagaagccaatcatgaaatgaactcccatcacaattgcttcaagaagaataaa 963
Db 577 CAACAGAGAGCCCAATCATGAGTCAATTCCCATTCACAAATTGCTTCAAGAGANTAAAA 636
QY 964 tacctaaggaatccaacttacaaggatgtgaaggaaccttcaaggagaactacaaacca 1023
Db 637 TACCTAGGAATCCAACCTACAGAGATGTGAAGGACCTCTTAAGAGAGAACTACAAACCA 696
QY 1024 ctgctcaatgnaataaagaagatacaacaataatggaagaacatccatgctcatgata 1083
Db 697 CTGCTTAATGAATTAACAGAGATACAAACAATGGAAGAAACATTCCATGCTCATGGGTA 756
QY 1084 gga 1086
Db 757 GGA 759
```

Search completed: March 30, 2002, 09:32:04
Job time: 13981 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 13:05:03 ; Search time 519.52 Seconds
(without alignments)
7539.881 Million cell updates/sec

Title: US-09-867-034-3

Perfect score: 4569
Sequence: 1 atgagtaaatctctcaacac.....tcgacgtctgcaatagcctag 4569

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_1101.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2492.2	54.5	3311	22	AA129502
2	2491.2	54.5	2854	22	AAH34879
3	2491.2	54.5	2854	22	AAH31787
4	2489.2	54.5	3111	20	AAZ09840
5	2485.6	54.4	2825	22	AAH46124
6	2477	54.2	2867	22	AAH33285
7	2473.6	54.1	2745	20	AAH31927
8	2469	54.0	2742	22	AAH46102
9	2467.6	54.0	3109	22	AAH35019
10	1517.6	33.2	2739	22	AAH46101
11	1517.6	33.2	2843	22	AAH46120

12	1512.8	33.1	2931	20	AAH81925	Murine ICAC-1 nuc
13	1129.8	24.7	3265	21	AAZ65095	Membrane-bound pro
14	1129.8	24.7	3265	22	AAH20992	Human PRO1124 cDNA
15	1129.8	24.7	3265	22	AAH44241	Human PRO1124 (UNC)
16	1015.2	22.2	4709	21	AAH69112	Human secreted pro
17	1012.4	22.2	5065	21	AAH69111	Human secreted pro
18	1012.4	22.2	5065	22	AAH69111	Human secreted pro
19	995.8	21.8	32042	20	AAZ09252	Human secreted pro
20	995.8	21.8	32042	20	AAH30011	Human CARD-4 DNA.
21	992.6	21.7	49999	20	AAZ23901	Human CARD-4 gene.
22	984.6	21.5	32199	22	AAH57673	Human LOBO homolog
23	981.4	21.5	50000	22	AAH54867	Human colorectal c
24	981.4	21.5	81145	22	AAH54868	Nucleotide sequenc
25	970.2	21.2	40328	21	AAH23815	Genomic nucleotide
26	969.4	21.2	14460	21	AAH23815	Human DAZ genomic
27	964.6	21.1	7207	22	AAH3404	Olfactory receptor
28	957.6	21.0	26190	22	AAH3279	Human kidney relat
29	957.4	21.0	18596	22	AAH31109	Human kidney relat
30	957.4	21.0	18596	22	AAH31215	Thymidylate syntha
31	940.6	20.6	23855	22	AAH3761	Human thymidylate
32	939.2	20.6	11520	22	AAH88291	Human kidney relat
33	937	20.5	68940	20	AAH37351	Human EST-derived
34	926.2	20.3	2616	21	AAH4335	Human Chromosome 6
35	919.6	20.1	41599	21	AAH23531	Clone 251688 of a
36	905	19.8	9339	21	AAH20962	Cosmid including s
37	905	19.8	9339	21	AAH34840	Human low adenosin
38	905	19.8	22421	21	AAH20966	Human adenosine re
39	905	19.8	22421	21	AAH3484	Human low adenosin
40	904.2	19.8	110000	22	AAH4800	Human adenosine re
41	888	19.4	5186	22	AAH88265	Nucleotide sequenc
42	839.8	18.4	138169	21	AAH34791	Human EST-derived
43	839.8	18.4	141589	21	AAH20913	Human adenosine re
44	839.8	18.4	141589	21	AAH21137	Human ELAM-1 polyn
45	839.8	18.4	141589	21	AAH21152	Human low adenosin

ALIGNMENTS

RESULT 1	
ID	AA129502
AA129502	standard; cDNA: 3311 BP.
AC	AA129502;
DT	12-OCT-2001 (first entry)
DE	C902P determined cDNA sequence.
XX	
KW	Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW	immunogenic; gene therapy; vaccine; colonic cancer; ss.
XX	
OS	Homo sapiens.
XX	
XX	WO200149716-A2.
PD	12-JUL-2001.
XX	
PF	29-DEC-2000; 2000OWO-US35596.
XX	
PR	30-DEC-1999; 99US-0476296.
PR	10-JAN-2000; 2000US-0480321.
PR	15-FEB-2000; 2000US-0504629.
PR	06-MAR-2000; 2000US-0519444.
PR	19-MAY-2000; 2000US-0575251.
PR	29-JUN-2000; 2000US-0609448.
PR	28-AUG-2000; 2000US-0649811.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
XX	King GE, Wang T, Jiang Y;
XX	

QY 3155 aactagaagagctgtccaaaatgacagagggtttacagacatatgcttcagatcaagttc 3214
|||||
Db 1682 aactagaagagctgtccaaaatgacagagggtttacagacatatgcttcagatcaagttc 1741
QY 3215 agacaatgagcttcattgattgcttttgggaccttcacaggaatggaagctgtcttc 3274
|||||
Db 1742 agacaatgagcttcattgattgcttttgggaccttcacaggaatggaagctgtcttc 1801
QY 3275 agcgtccatccagcttgaagagtaaggatcaacctccagaacagcagtgatgaatg 3334
|||||
Db 1802 agcgtccatccagcttgaagagtaaggatcaacctccagaacagcagtgatgaatg 1861
QY 3335 gcaacagtatcgttgacagcagccgtgggaaggagcaactttttcttaccctggacaa 3394
|||||
Db 1862 gcaacagtatcgttgacagcagccgtgggaaggagcaactttttcttaccctggacaa 1921
QY 3395 tgcagctccccaatctctctcttggaatccagtgacagaagaagtggtctttag 3454
|||||
Db 1922 tgcagctccccaatctctctcttggaatccagtgacagaagaagtggtctttag 1981
QY 3455 tggacaaaacacaaaatgagctaccctccaaatccagcaattgctaagttggacctt 3514
|||||
Db 1982 tggacaaaacacaaaatgagctaccctccaaatccagcaattgctaagttggacctt 2041
QY 3515 ggaatatagctgtgaagcagctcacaaaccttgaccttgacttgacgtcccgctgct 3574
|||||
Db 2042 ggaatatagctgtgaagcagctcacaaaccttgaccttgacttgacgtcccgctgct 2101
QY 3575 ccaatgctaccctgctccccaatcaacagtgacttccaaaacagaacagcagcaaat 3634
|||||
Db 2102 ccaatgctaccctgctccccaatcaacagtgacttccaaaacagaacagcagcaaat 2161
QY 3635 tcccaagcctctctgtttagttatgcaaatatcgcgaagagcctcccaattctcagg 3694
|||||
Db 2162 tcccaagcctctctgtttagttatgcaaatatcgcgaagagcctcccaattctcagg 2221
QY 3695 ccagtgacagcagcctgtgtaaatcagtgaaatggaacagcttacccttgaaactcagg 3754
|||||
Db 2222 ccagtgacagcagcctgtgtaaatcagtgaaatggaacagcttacccttgaaactcagg 2281
QY 3755 ataattgagcagtgctgactaactaagatgacgctgcttactcaagatattcacaa 3814
|||||
Db 2282 ataattgagcagtgctgactaactaagatgacgctgcttactcaagatattcacaa 2341
QY 3815 ctatgacacgaatggtatgatacagtgtaaaaagtgcggcctctggagaggttaacgag 3874
|||||
Db 2342 ctatgacacgaatggtatgatacagtgtaaaaagtgcggcctctggagaggttaacgag 2401
QY 3875 ccagacgagagtgatgatacccaagcagtgagcactgtacatcctggctgagattgaga 3934
|||||
Db 2402 ccagacgagagtgatgatacccaagcagtgagcactgtacatcctggctgagattgaga 2461
QY 3935 atgatgaaatacaatggaatccaccaagaacctgaataataagaatgatttcaacaa 3994
|||||
Db 2462 atgatgaaatacaatggaatccaccaagaacctgaataataagaatgatttcaacaa 2521
QY 3995 agcaagtggttcaagcagaacatctcggagaggtcattgttgcttctgattgccaa 4054
|||||
Db 2522 agcaagtggttcaagcagaacatctcggagaggtcattgttgcttctgattgccaa 2581
QY 4055 atgtcccataccctgattcttcccaactggccaataccacagcctggaaggagaaattc 4114
|||||
Db 2582 atgtcccataccctgattcttcccaactggccaataccacagcctggaaggagaaattc 2641
QY 4115 acgggggagcttcataatcactgacttgacagcctcctgggagatgattatgacattgaa 4174
|||||
Db 2642 acgggggagcttcataatcactgacttgacagcctcctgggagatgattatgacattgaa 2701
QY 4175 cagctcaaatatatacttctgaaataagatacagatattcttgattctcagagacagttca 4234
|||||
Db 2702 cagctcaaatatatacttctgaaataagatacagatattcttgattctcagagacagttca 2761

QY 4235 atgaatctctcaagtgataactactgtctctcaatcccaagggaagccaacttgaggag 4294
|||||
Db 2762 atgaatctctcaagtgataactactgtctctcaatcccaagggaagccaacttgaggag 2821
QY 4295 tcttttgtttaaaccaaaaacatlaacttttgaataatgacagatctttcattgcta 4354
|||||
Db 2822 tcttttgtttaaaccaaaaacatlaacttttgaataatgacagatctttcattgcta 2881
QY 4355 ttcaagcgtgtgataaggtgatctgaaatcagaataatccaacttgcagagatctt 4414
|||||
Db 2882 ttcaagcgtgtgataaggtgatctgaaatcagaataatccaacttgcagagatctt 2941
QY 4415 tgttatctccccaagagctccgccaagagacacccagctcctgataaagctctgctctt 4474
|||||
Db 2942 tgttatctccccaagagctccgccaagagacacccagctcctgataaagctctgctctt 3001
QY 4475 gtccataatcataltaacacagacacatctcctgacatcattttaaattatgtga 4534
|||||
Db 3002 gtccataatcataltaacacagacacatctcctgacatcattttaaattatgtga 3061
QY 4535 agtgatagagaactgacgtgtgtcaatagcctag 4569
|||||
Db 3062 agtgatagagaactgacgtgtgtcaatagcctag 3096

RESULT 2
AAH34879
ID AAH34879 standard; cDNA: 2854 BP.
XX
AC AAH34879:
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1961.
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1961.
XX
KW colorectal carcinoma; chromosome 1; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PE 28-SEP-2000; 2000MO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
XX
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
XX
DR P-PSDB; AAG75474.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 3462-3463; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell

QY 3524 gtctgcaagcaagctcaacaaaccttgaccctgactgtcaagctccgctgctccaatgtcta 3583
Db 1734 gtctgcaagcaagctcaacaaaccttgaccctgactgtcaagctccgctgctccaatgtcta 1793
QY 3584 cccctgctcccaattacagtgagcttccaaagcaagcaagcaagcaagcaagcaagcaagcc 3643
Db 1794 cccctgctcccaattacagtgagcttccaaagcaagcaagcaagcaagcaagcaagcc 1853
QY 3644 ctctggtgagtttatgcaaatatttcgcaagagagctcccccattcttcaggccagtgca 3703
Db 1854 ctctggtgagtttatgcaaatatttcgcaagagagctcccccattcttcaggccagtgca 1913
QY 3704 cagcccttattgaatcaatgagtaagtaaaacagttacttgaactactgagtaagag 3763
Db 1914 cagcccttattgaatcaatgagtaagtaaaacagttacttgaactactgagtaagag 1973
QY 3764 caggtgtgactgactactaagagatgacggtgtctactcaaggtatttccaaactatgaca 3823
Db 1974 caggtgtgactgactactaagagatgacggtgtctactcaaggtatttccaaactatgaca 2033
QY 3824 cgaatgtgagatgacagtgtaaaagtgcggtcttggagagaggttaacgcagccagcgga 3883
Db 2034 cgaatgtgagatgacagtgtaaaagtgcggtcttggagagaggttaacgcagccagcgga 2093
QY 3884 gagtgaatcccccaagagagtgagacgactgacatacctcggtcggttgaagatgtgaaa 3943
Db 2094 gagtgaatcccccaagagagtgagacgactgacatacctcggtcggttgaagatgtgaaa 2153
QY 3944 tacaatggaatcccaacgaagcctgaaatlaaagatgagtltcaacaacgaagcagtgct 4003
Db 2154 tacaatggaatcccaacgaagcctgaaatlaaagatgagtltcaacaacgaagcagtgct 2213
QY 4004 gtctcagaagaacatccctcggagaggtctatttggcttctgagtcccaaatgtgtccca 4063
Db 2214 gtctcagaagaacatccctcggagaggtctatttggcttctgagtcccaaatgtgtccca 2273
QY 4064 tactgactcttccacactgagcaaatccacgacagctgaagcggaatttccaggggca 4123
Db 2274 tactgactcttccacactgagcaaatccacgacagctgaagcggaatttccaggggca 2333
QY 4124 gtctcaatcaatcgacttgagcagctcctgggagatgattatgacatgagcaagctcaca 4183
Db 2334 gtctcaatcaatcgacttgagcagctcctgggagatgattatgacatgagcaagctcaca 2393
QY 4184 agtatataatcgaataagataagatattcttgatctcagagcaagttcaatgatactc 4243
Db 2394 agtatataatcgaataagataagatattcttgatctcagagcaagttcaatgatactc 2453
QY 4244 ttcaagtaataactactgtctcattcccaagaagcaacactgaggaagcttttgt 4303
Db 2454 ttcaagtaataactactgtctcattcccaagaagcaacactgaggaagcttttgt 2513
QY 4304 ttaaaccgaagaacattacttggaaatgagacagatctttcatctgcatcagagctg 4363
Db 2514 ttaaaccgaagaacattacttggaaatgagacagatctttcatctgcatcagagctg 2573
QY 4364 ttgtaagtgatgacttggaatcagaataatcacaatgacagagatatttggattatc 4423
Db 2574 ttgtaagtgatgacttggaatcagaataatcacaatgacagagatatttggattatc 2633
QY 4424 ctcccaacacccgcaagagacacttctgtagaagcgtctgctccttccataa 4483
Db 2634 ctcccaacacccgcaagagacacttctgtagaagcgtctgctccttccataa 2693
QY 4484 ttcatatacaacagcaacttccctggcattcacattttaaanaatitgtagaagtgatag 4543
Db 2694 ttcatatacaacagcaacttccctggcattcacattttaaanaatitgtagaagtgatag 2753
QY 4544 gagaactgacgtgcatagcttag 4569
Db 2754 gagaactgacgtgcatagcttag 2779

RESULT 3
AA81787
ID AA81787 standard; cDNA: 2854 BP.
XX
AC AA81787;
XX
DT 12-JUN-2001 (first entry)
XX
DE Human secreted protein gene 1 SEQ ID NO:11.
XX
KW Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; anti-inflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
KW antialzheimer; antiparkinsonian; antimicrobial; vulnerary; gene therapy;
KW immune disorder; hyperproliferative disorder; cardiovascular disease;
KW cancer; angiogenic disorder; neurological disorder; infectious disease;
KW wound healing; regeneration; chemotaxis; chromosome 1; ss.
XX
OS Homo sapiens.
XX
PN WO200112775-A2.
XX
PD 22-FEB-2001.
XX
PE 16-AUG-2000; 2000WO-US22325.
XX
PR 17-AUG-1999; 99US-0149182.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;
PI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;
XX
DR WP1: 2001-147550/15.
XX
P-PSDB: AAB74733.
XX
PT Nucleic acids encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Claim 1: Page 441: 485bp; English.
XX
CC AA81787 to AA81817 encode the human secreted proteins given in AAB74733
CC to AAB74772. Human secreted proteins can have activities based on the
CC tissues and cells they are expressed in. Example of activities include:
CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;
CC antinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;
CC anticonvulsant; antialzheimer; antiparkinsonian; antimicrobial; and
CC vulnerary. Human secreted proteins can be used in gene therapy and
CC vaccine. Human secreted protein nucleotide sequences (NAMI) and proteins
CC (PEP1) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, NAMI
CC and PEP1 may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patients genome
CC that affect the activity of proteins by expressing inactive proteins or
CC to supplement the patients own production of polypeptides. Disorders that
CC may be prevented, diagnosed and/or treated include immune disorders,
CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
CC angiogenic disorders, neurological disorders, infectious diseases and/or
CC for promoting wound healing, regeneration and/or chemotaxis. AA81778 to
CC AA81786 and AA81732 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 2854 BP; 885 A; 628 C; 643 G; 698 T; 0 other;

Query Match 54.5%; Score 2491.2; DB 22; Length 2854;
Best Local Similarity 91.5%; Pred. No. 0;
Matches 2768; Conservative 0; Mismatches 3; Indels 255; Gaps 2;

OY	1544	gggaaatacaaggagagatgtaacgcaatggggccatttaaggtctgtgttcattctgta	1603
Db	9	ggagaaatacaaggagagatgtaacgcaatggggccatttaaggtctgtgttcattctgta	68
OY	1604	ttcttcacactctctgaaggggccctctgtgaattcactcattcagctgtaacaaatctgct	1663
Db	69	ttcttcacactctctgaaggggccctctgtgaattcactcattcagctgtaacaaatctgct	128
OY	1664	atgaagcattgtcgtctgtgcaatcgaccacaatgtgcccagaatgtgaanaactcattcaac	1723
Db	129	atgaagcattgtcgtctgtgcaatcgaccacaatgtgcccagaatgtgaanaactcattcaac	188
OY	1724	aaataaagggggaattacacgttcacaaagtatgaagaaggaggttcaggagaactctctc	1783
Db	189	aaata-----aaata-----	193
OY	1784	ttcccccgtcaataatacatacacacacacacacagcacagaactcgtgtgcacacacaca	1843
Db	194	-----	193
OY	1844	cgcccatgcaacacacagcagacatacacacacacacacgcaacgcttcagaaggacatgtgaccc	1903
Db	194	-----aagacatggtgacc	209
OY	1904	aggcatctctgtatctcgtcttggaagctcacaggaagcgaatttatctcaaaatgtgtcca	1963
Db	210	aggcatctctgtatctctgttggaagctcacaggaagcgaatttatctcaaaatgtgtcca	269
OY	1964	ttttgatctccgtaaacatgtgaagaacaaaggctaacatgttgagacccaacattgtgacct	2023
Db	270	ttttgatctccgtaaacatgtgaagaacaaaggctaacatgttgagacccaacattgtgacct	329
OY	2024	acaaaatgctgatagtctctcgtgtgctgagttcattctctcccaagttatgtgaaccttaca	2083
Db	330	acaaaatgctgatagtctctcgtgtgagttcattctctcccaagttatgtgaaccttaca	389
OY	2084	ctggagcagatgtgggaactgtgggagagaaagggtgaagaagttccactcactcctgatttca	2143
Db	390	ctggagcagatgtgggaactgtgggagagaaagggtgaagaagttccactcactcctgatttca	449
OY	2144	ttgcagggaaaaaagttagctgtaatatgagaccaaagtagggcattgtgtccatgtggtggg	2203
Db	450	ttgcagggaaaaaagttagctgtaatatgagaccaaagtagggcattgtgtccatgtggtggg	509
OY	2204	ctcatctcagatggggaggtatttgaacgagtacataatgatgatgaagaattctactatcca	2263
Db	510	ctcatctcagatggggaggtatttgaacgagtacataatgatgatgaagaattctactatcca	569
OY	2264	atggaagaatatcaagcagctaaagatgttcagcaggttatctcgtgtccaaatgtgtaaaga	2323
Db	570	atggaagaatatcaagcagctaaagatgttcagcaggttatctcgtgtccaaatgtgtaaaga	629
OY	2324	agttctcagggggagcagctgtttacacaaaagaatgtgacattcaataagttaacagagctct	2383
Db	630	agttctcagggggagcagctgtttacacaaaagaatgtgacattcaataagttaacagagctct	689
OY	2384	atgaaaaaaggatgtgagtttgtgtctccaalcccgcgcagacggaagaaggtctctataatgt	2443
Db	690	atgaaaaaaggatgtgagtttgtgtctccaalcccgcgcagacggaagaaggtctctataatgt	749
OY	2444	ttgcacaacaatgtgtattctataatgttggaattctctgtacagacaacacacacaagaag	2503
Db	750	ttgcacaacaatgtgtattctataatgttggaattctctgtacagacaacacacacaagaag	809
OY	2504	cttcaaacacagcaaaaatcaaaaatgtgaactctccgaagcagatgggaatgtgtccgtgatt	2563
Db	810	cttcaaacacagcaaaaatcaaaaatgtgaactctccgaagcagatgggaatgtgtccgtgatt	869
OY	2564	cttggagctttaaagaaaaaacctccttatgtacaacaacagccacaaatcccaacttccat	2623
Db	870	cttggagctttaaagaaaaaacctccttatgtacaacaacagccacaaatcccaacttccat	929
OY	2624	tgtgtcagattgtgaacaagaattgtgtgttttagtcccttgacaacatctggaagcatgtgcga	2683

Db	930	tgctgagatcgtggacaagaatctgtgtgtctttagtccctctggacaactctggaaagcagtgctga	989
Qy	2684	ctgtgtaacgcctcccaatctgcacgtgaatccaagcagcgccagctcttctctgtcgtgcacagtgtg	2743
Db	990	ctgtgtaacgcctccaatctgcacgtgaatccaagcagcgccagctcttctctgtcgtgcacagtggtg	1049
Qy	2744	agctggggtctctgtggttgggaatgtgacaattgcacagtgtcgtcccatgyltacaaaaatgaac	2803
Db	1050	agctggggtctctgtggttgggaatgtgacaattgcacagtgtcgtcccatgyltacaaaaatgaa	1109
Qy	2804	tcatacagttaaacagctgtgcacgttgacatggaggaacacactctgcacaaagattactctgacgcag	2863
Db	1110	tcatacagttaaacagctgtgcacgttgacatggaggaacacactctgcacaaagattactctgacgcag	1169
Qy	2864	cttccagaagggagcgtcccaactctgcacagcggtcttcgatctgcacatttaagtatgtgcacac	2923
Db	1170	cttccagaagggagcgtcccaactctgcacagcggtcttcgatctgcacatttaagtatgtgcacac	1216
Qy	2924	atttgccgttcttccatgacacacagcagttatgtggagtgatgcgacaagaaaaatcccaaat	2983
Db	1217	-----	1216
Qy	2984	ggggcctctctggccctgcgcagcttaagtataggaaagaatatcccaactgatatgtgaa	3043
Db	1217	-----ggtataggaaagaatatcccaactgatatgtgaa	1253
Qy	3044	ttgtcgtctgtcaggaatctggggagaagacaacacatlaagtgtgggtcttlaacggggccaac	3103
Db	1254	ttgtcgtctgtcaggaatctggggagaagacaacacatlaagtgtgggtcttlaacggggccaac	1313
Qy	3104	aaagtgtgtgccatccacacacagtctgtcttggggccctctgcacgtccaaagactagaag	3163
Db	1314	aaagtgtgtgccatccacacacagtctgtcttggggccctctgcacgtccaaagactagaag	1373
Qy	3164	agctgtccaaaaatgacagaaggtttacacgaactatgtcttcagatccaagtccaagaacatg	3223
Db	1374	agctgtccaaaaatgacagaaggtttacacgaactatgtcttcagatccaagtccaagaacatg	1433
Qy	3224	gacctatgatctcttctggggcccttctacacaggaatagaagctgtctccacgcgtcca	3283
Db	1434	gacctatgatctcttctggggcccttctacacaggaatagaagctgtctccacgcgtcca	1493
Qy	3284	tccacgtctgagagtaaggattaaacctccctccagaacacagccagtgtgaatgagcacagtga	3343
Db	1494	tccacgtctgagagtaaggattaaacctccctccagaacacagccagtgtgaatgagcacagtga	1553
Qy	3344	tcgtgtgacagcacgtgtggaaaagagcaacttgttctctatcaactcgtgcacaatgtcagctc	3403
Db	1554	tcgtgtgacagcacgtgtggaaaagagcaacttgttctctatcaactcgtgcacaatgtcagctc	1613
Qy	3404	cccacaatccctctctgggaatcccaatgtagacaagaacaggtgtgcttttagtggacaaca	3463
Db	1614	cccacaatccctctctgggaatcccaatgtagacaagaacaggtgtgcttttagtggacaaca	1673
Qy	3464	acacacaaaatgtgctactcccaatcccaatcccaaggaatgtcacaaggtltgacacttggaaataca	3523
Db	1674	acacacaaaatgtgctactcccaatcccaatcccaaggaatgtcacaaggtltgacacttggaaataca	1733
Qy	3524	gtctgtcaagcnaagctccaacacacttgcacccctgcagctgtcaagtcaccgtgtgcgttccaatgtcta	3583
Db	1734	gtctgtcaagcnaagctccaacacacttgcacccctgcagctgtcaagtcaccgtgtgcgttccaatgtcta	1793
Qy	3584	cccctgtccctccaattacacatgacttcccaaaacgaagaagaacacagcnaaatctcccacac	3643
Db	1794	cccctgtccctccaattacacatgacttcccaaaacgaagaagaacacagcnaaatctcccacac	1853
Qy	3644	ctctgtgatttatatcaaatatctgcacaaagagcctccccaattctcagggccaagtgtca	3703
Db	1854	ctctgtgatttatatcaaatatctgcacaaagagcctccccaattctcagggccaagtgtca	1913
Qy	3704	cagccctgattgaaatcagatgaatggaaaaacagttaccttggaaactatgtgataatggag	3763

```

Db 1914 cagccctgattgaaatcagtgaaatgaaacagtacccttggaactactgataatgag 1973
OY 3764 caggtgctgactgactaagaatgacggtgtctactcaaggtatttcaactatgaca 3823
Db 1974 caggtgctgactgactaagaatgacggtgtctactcaaggtatttcaactatgaca 2033
OY 3824 cgaatggtagaataagtgaaagtgccgggtcttgaggaggttaacgcaacgaacgca 3883
Db 2034 cgaatggtagaataagtgaaagtgccgggtcttgaggaggttaacgcaacgaacgca 2093
OY 3884 gagtgtataccccagagagtgagacactgtacataccgtgctgattgagaatgataa 3943
Db 2094 gagtgtataccccagagagtgagacactgtacataccgtgctgattgagaatgataa 2153
OY 3944 tacatggaatcccaacgaactgaaataaagatgagtgatccaacgaacgaaggt 4003
Db 2154 tacatggaatcccaacgaactgaaataaagatgagtgatccaacgaacgaaggt 2213
OY 4004 gtttcagcagaacatcctcggagggtcattgtggtcttcgtatgcccgaatgtccca 4063
Db 2214 gtttcagcagaacatcctcggagggtcattgtggtcttcgtatgcccgaatgtccca 2273
OY 4064 tacctgactcttcaccactgagcaaatcacccgactgaagcggaataatcaaggggca 4123
Db 2274 tacctgactcttcaccactgagcaaatcacccgactgaagcggaataatcaaggggca 2333
OY 4124 gtctcatataactgactgagcagctcctgsgggtatgattgacatgtaacagctaca 4183
Db 2334 gtctcatataactgactgagcagctcctgsgggtatgattgacatgtaacagctaca 2393
OY 4184 agtatatcatcgaataagtaagtaacttctgtatctcagagacaagtccaatgatactc 4243
Db 2394 agtatatcatcgaataagtaagtaacttctgtatctcagagacaagtccaatgatactc 2453
OY 4244 ttcaagtgtaaactactgtcttcacccaagaaggaagcaactctgaggaagctctttgt 4303
Db 2454 ttcaagtgtaaactactgtcttcacccaagaaggaagcaactctgaggaagctctttgt 2513
OY 4304 ttaaacccgaaaatacttcttgaagaatgacacagatcttcaattgcttcaagctg 4363
Db 2514 ttaaacccgaaaatacttcttgaagaatgacacagatcttcaattgcttcaagctg 2573
OY 4364 ttgataagtgatctgaaatcagaataatcacaactgtcagagtaacttctgtattac 4423
Db 2574 ttgataagtgatctgaaatcagaataatcacaactgtcagagtaacttctgtattac 2633
OY 4424 ctccacagactccgcagagacacactagtcctgtatgaaagctgtgcttctgtccata 4483
Db 2634 ctccacagactccgcagagacacactagtcctgtatgaaagctgtgcttctgtccata 2693
OY 4484 ttcatatacagaacacatctcctgcatcattttaaataatgagaagtgatag 4543
Db 2694 ttcatatacagaacacatctcctgcatcattttaaataatgagaagtgatag 2753
OY 4544 gagaactgcaagctgtaataagcctag 4569
Db 2754 gagaactgcaagctgtaataagcctag 2779

```

```

RESULT 4
AAZ09840
ID AAZ09840 standard; cDNA; 3111 BP.
XX
AC AAZ09840;
XX
DT 26-NOV-1999 (first entry)
DE
XX Human membrane spanning protein MSP-5 cDNA fragment 2.
XX
KW Membrane spanning protein; MSP; human; treatment; diagnosis; prevention;
KW neoplastic disorder; immunological disorder; reproductive disorder.
KW MSP-5; ds.
XX

```

```

OS Homo sapiens.
XX
PN WO946380-A2.
XX
PD 16-SEP-1999.
XX
PF 09-MAR-1999; 99WO-US05073.
XX
PR 13-MAR-1998; 98US-0039064.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Bandman O, Lal P, Hillman JL, Yue H, Corley NC;
PI Guegler KJ, Kaser MR, Baughn MR, Shah P;
XX
DR WPI: 1999-551409/46.
DR P-PSDB: AAY33298.
XX
PT New human membrane spanning proteins used to, e.g. prevent and treat
PT neoplastic disorders -
XX
PS Example 1: Page 80-81; 81pp; English.
XX
CC This invention describes novel human membrane spanning proteins (MSPs),
CC and the polynucleotides encoding them. The products of the invention are
CC used to diagnose, prevent and treat neoplastic, immunological and
CC reproductive disorders. This sequence encodes a human membrane spanning
CC protein MSP-5 fragment.
XX
SQ Sequence 3111 BP; 968 A; 664 C; 691 G; 788 T; 0 other;

```

Query Match 54.5%; Score 2489.2; DB 20; Length 3111;

Best Local Similarity 91.5%; Pred. No. 0; Mismatches 3; Indels 255; Gaps 2;

Matches 2766; Conservative 0; Mismatches 3; Indels 255; Gaps 2;

```

OY 1546 gaatacagaaggagatgtacagcaatggggccattaaagttctgttcaactgatt 1605
Db 10 gaatacagaaggagatgtacagcaatggggccattaaagttctgttcaactgatt 69
OY 1606 ctccactctcagaaggggccctgagtaattcaactcactcagctgaaacaaatgctat 1665
Db 70 ctccactctcagaaggggccctgagtaattcaactcactcagctgaaacaaatgctat 129
OY 1666 gaagcattgtcgttgcgaatgaccccaatgtgcagaagaatgaaacctatcaaca 1725
Db 130 gaagcattgtcgttgcgaatgaccccaatgtgcagaagaatgaaacctatcaaca 189
OY 1726 ataaagggggagtacagctcaagaatgaggaaggagatcagagagaactctctt 1785
Db 190 ata----- 192
OY 1786 ccccgctcaaatatatacatcacacacacacacagcagaagctcgtgtgacacacacag 1845
Db 193 ----- 192
OY 1846 cccatgcacacagcagacatacaacgacacacagcagctcagaaagacatgtgaccag 1905
Db 193 -----aaggaatggtgaccag 210
OY 1906 gcatcctgtatctgtctgaaactacaggaagcatttatttcaaaaatgttccatt 1965
Db 211 gcatcctgtatctgtctgaaactacaggaagcatttatttcaaaaatgttccatt 270
OY 1966 ttgattcctgaaacatggaagacaaaggctgactatgtgagaccacaaacttgagacctac 2025
Db 271 ttgattcctgaaacatggaagacaaaggctgactatgtgagaccacaaacttgagacctac 330
OY 2026 aaaaatgtctatgttctgtgtgtgagctactcctccaggttaatgagaacctaacct 2085
Db 331 aaaaatgtctatgttctgtgtgtgagctactcctccaggttaatgagaacctaacct 390
OY 2086 gagcagatggcgaactgtgagagagaagggtgaagaatccacactcactccgatttcaat 2145

```

```
|||||
Db gagcagatgaggcaactctgagagaagggtgaaggaatccaccctcactccgattcaatt 450
2146 gcaagaaaaaagattagctgaaatctggaaccaagaagttaggcatctgttccatgagtggtc 2205
|||||
Db gcaagaaaaaagattagctgaaatctggaaccaagaagttaggcatctgttccatgagtggtc 510
2206 catctacgattgaggagatttgcagagtaacaataatgataagaaattctacttaccat 2265
|||||
Db catctacgattgaggagatttgcagagtaacaataatgataagaaattctacttaccat 570
2266 ggaagaatacaagcagtaagaatgttccagcaggtattactcgtgtacaaatgtgtataagaag 2325
|||||
Db ggaagaatacaagcagtaagaatgttccagcaggtattactcgtgtacaaatgtgtataagaag 630
571 ggaagaatacaagcagtaagaatgttccagcaggtattactcgtgtacaaatgtgtataagaag 630
2326 tgcacaggaagcagcgtttaaaccaaaagatgcacattcaataaagttaacagagctctat 2385
|||||
Db tgcacaggaagcagcgtttaaaccaaaagatgcacattcaataaagttaacagagctctat 690
631 tgcacaggaagcagcgtttaaaccaaaagatgcacattcaataaagttaacagagctctat 690
2386 gaaaaaagatgtaggttctgtctccaaatcccgccagacggaagagctctataatgtt 2445
|||||
Db gaaaaaagatgtaggttctgtctccaaatcccgccagacggaagagctctataatgtt 750
691 gaaaaaagatgtaggttctgtctccaaatcccgccagacggaagagctctataatgtt 750
2446 gcaacaatgttgttctctatatagttgaattctgtctgaacaacaacccacaagaagagct 2505
|||||
Db gcaacaatgttgttctctatatagttgaattctgtctgaacaacaacccacaagaagagct 810
751 gcaacaatgttgttctctatatagttgaattctgtctgaacaacaacccacaagaagagct 810
2506 ccaacaagcaaaaatcaaaaatgcaatctccgaagcacaatgggaagtgcgtatctc 2565
|||||
Db ccaacaagcaaaaatcaaaaatgcaatctccgaagcacaatgggaagtgcgtatctc 870
811 ccaacaagcaaaaatcaaaaatgcaatctccgaagcacaatgggaagtgcgtatctc 870
2566 gaggaacttaagaaaacactcctctatgacaacaacagccaccaatcccaacttctcatg 2625
|||||
Db gaggaacttaagaaaacactcctctatgacaacaacagccaccaatcccaacttctcatg 930
871 gaggaacttaagaaaacactcctctatgacaacaacagccaccaatcccaacttctcatg 930
2626 ctgcaagtttgacaaaagaattgtgtgtttagtcccttgacaaaatctggaagatgagcact 2685
|||||
Db ctgcaagtttgacaaaagaattgtgtgtttagtcccttgacaaaatctggaagatgagcact 990
931 ctgcaagtttgacaaaagaattgtgtgtttagtcccttgacaaaatctggaagatgagcact 990
2686 gtaaacgcctcaatctcagatgaatcaagcagccagccttctccctgcgcagacgctgag 2745
|||||
Db gtaaacgcctcaatctcagatgaatcaagcagccagccttctccctgcgcagacgctgag 1080
991 gtaaacgcctcaatctcagatgaatcaagcagccagccttctccctgcgcagacgctgag 1080
2746 ctggggtctctgggttgagatggtgacatttgacagtgctgcgccatgtaacaaatgaaatc 2805
|||||
Db ctggggtctctgggttgagatggtgacatttgacagtgctgcgccatgtaacaaatgaaatc 1110
1051 ctggggtctctgggttgagatggtgacatttgacagtgctgcgccatgtaacaaatgaaatc 1110
2806 atcacagtaaacagtgagtgagtgacagagacacactgcgcaaaaagattacctgcgcagcct 2865
|||||
Db atcacagtaaacagtgagtgagtgagtgacagagacacactgcgcaaaaagattacctgcgcagcct 1170
1111 atcacagtaaacagtgagtgagtgagtgacagagacacactgcgcaaaaagattacctgcgcagcct 1170
2866 tcaaggaagacgttccatctgcagcgggtctgcagtcgcatcttaactatgtgcaaat 2925
|||||
Db tcaaggaagacgttccatctgcagcgggtctgcagtcgcatcttaactatgtgcaaat 1225
1171 tcaaggaagacgttccatctgcagcgggtctgcagtcgcatcttaactatgtgcaaat 1225
2926 tgcgctgttctcatgacacacagcagttatgaggagtgcgacaagaanaatcccaattgg 2985
|||||
Db tgcgctgttctcatgacacacagcagttatgaggagtgcgacaagaanaatcccaattgg 1215
1216 ----- 1215
2986 ggcctcttgccctgcagccttagtgattagaagaanaatatcccaatgtagtctgaaatc 3045
|||||
Db ggcctcttgccctgcagccttagtgattagaagaanaatatcccaatgtagtctgaaatc 1254
1216 -----gtagttaggaagaanaatatcccaatgtagtctgaaatc 1254
3046 gtcgctgtgacgagtgagggaagaacaacataagtggtgctttaaagcggttaacaa 3105
|||||
Db gtcgctgtgacgagtgagggaagaacaacataagtggtgctttaaagcggttaacaa 1314
1255 gtcgctgtgacgagtgagggaagaacaacataagtggtgctttaaagcggttaacaa 1314
3106 agtgggtgcatcatcatcacacagtcgcttgaggccctctgcagctcaagaactagaggag 3165
|||||
Db agtgggtgcatcatcatcacacagtcgcttgaggccctctgcagctcaagaactagaggag 1374
1315 agtgggtgcatcatcatcacacagtcgcttgaggccctctgcagctcaagaactagaggag 1374
3166 ctgttccaaaatgacaggaggttaccagacatagtctccagatcaagtctcagaacaatggc 3225
|||||
```


Dh 742 acaaatgttattctatagttaattcttctgtacagaacaaacacaaagaagctcc 801
QY 2508 aaacaagcaaaaatcaaatatgcaatcttcgcagacacatgaggagtgatccgttgattctga 2567
Dh 802 aaacaagcaaaaatcaaaaatgcaatcttcgcagacacatgaggagtgatccgttgattctga 861
QY 2568 ggaatttaagaanaaacccactctatgacaacacagccacaaatccacacttctcatgtct 2627
Dh 862 ggaatttaagaanaaacccactctatgacaacacagccacaaatccacacttctcatgtct 921
QY 2628 gcaagattggaacaaagaatttggtttttaagttcttgacaacacttggaagcatggagcagctg 2687
Dh 922 gcaagattggaacaaagaatttggtttttaagttcttgacaacacttggaagcatggagcagctg 981
QY 2688 taacagcctcaatcgactgaaatcaacagcagcagcacttcttcgtctgcagacagcttgact 2747
Dh 982 taacgctcctaactcgactgaaatcaacagcagcagcacttcttcgtctgcagacagcttgact 1041
QY 2748 ggggtcctgggttggtgattgacatttgaaagtgcctgcaccaatgatacaaaatgaaatcctat 2807
Dh 1042 ggggtcctgggttggtgattgacatttgaaagtgcctgcaccaatgatacaaaatgaaatcctat 1101
QY 2808 acagataaacaagctgacagtgacaggaacacactgcgcaaaagaattacactgacagcagcttc 2867
Dh 1102 acagataaacaagctgacagtgacaggaacacactgcgcaaaagaattacactgacagcagcttc 1161
QY 2868 aggaaggagcgtlccactctgcagcagcgtctgcactgcgcaatttactgatalgtgacaaattt 2927
Dh 1162 aggaaggagcgtlccactctgcagcagcgtctgcactgcgcaatttactgatalgtgacaaattt 1204
QY 2928 ggcctgtttccatgacacacagcagttatgggaagtgcgacagaanaatcccaattgggc 2987
Dh 1205 ----- 1204
QY 2988 ctctcctggcctgacagcttaagtattgaagaanaatatccaactgatgatgtgaatttgt 3047
Dh 1205 -----ggaatttaaggaaagaataatccaactgataggatcttgaaatttgt 1245
QY 3048 gctctgacagatgagggaagaacacactataaagtgggttgctttaacagaggtcaacaacaag 3107
Dh 1246 gctctgacagatgagggaagaacacactataaagtgggttgctttaacagaggtcaacaacaag 1305
QY 3108 tgggtgcataatccacacagtcgcttttggggcctctgcagctcaagaataagtagagagct 3167
Dh 1306 tgggtgcataatccacacagtcgcttttggggcctctgcagctcaagaataagtagagagct 1365
QY 3168 gtccaaatatgacagagaggtttacagacatatgcttcagatcaaatgttaagaacaatgtgcct 3227
Dh 1366 gtccaaatatgacagagaggtttacagacatatgcttcagatcaaatgttaagaacaatgtgcct 1425
QY 3228 catgtatgcttttggggccttccatcaggaaatggaagctgtctctcagcgtccatcca 3287
Dh 1426 catgtatgcttttggggccttccatcaggaaatggaagctgtctctcagcgtccatcca 1485
QY 3288 gcttggaagttaaggatlaaacctccagaacagcagctgatatgaatggcacagtgtatcgt 3347
Dh 1486 gcttggaagttaaggatlaaacctccagaacagcagctgatatgaatggcacagtgtatcgt 1545
QY 3348 ggaagacacgttggaaagaagaacttggttcttatccacggcagcaatgacgctcccca 3407
Dh 1546 ggaagacacacgttggaaagaagaacttggttcttatccacggcagcaatgacgctcccca 1605
QY 3408 aatcctctctgagatcccgctgacagaagaagcagctgctttgattgagcaaaaacac 3467
Dh 1606 aatcctctctgagatcccgctgacagaagaagcagctgctttgattgagcaaaaacac 1665
QY 3468 caaatggtcctacctccaatcccaaggcatgttaagttggacactggaataacagct 3527
Dh 1666 caaatggtcctacctccaatcccaaggcatgttaagttggacactggaataacagct 1725
QY 3528 gcaagcagaagctcaaaaacacttgacacctgactgttcaagctcccggtgcacaaatgtacacct 3587
Dh 1726 gcaagcagaagctcaaaaacacttgacacctgactgttcaagctcccggtgcacaaatgtacacct 1785

QY 3588 gctccaattacagtgacttccaaaacgaaacagagacacagcaaatctcccaagcctct 3647
Dh 1786 gctccaattacagtgacttccaaaacgaaacagagacacagcaaatctcccaagcctct 1845
QY 3648 ggtatgtttatgcaaatatttgcgaaggagcctcccaattctcagggccagtgtaacagc 3707
Dh 1846 ggtatgtttatgcaaatatttgcgaaggagcctcccaattctcagggccagtgtaacagc 1905
QY 3708 cctgatgtcaatcagtgaaatgaaacacagttactcttgaaactaagtatgatgagcag 3767
Dh 1906 cctgatgtcaatcagtgaaatgaaacacagttactcttgaaactaagtatgatgagcag 1965
QY 3768 tgcgtatgacttaagaatgacagctgtctactcaaggtatttccaaactatgacacgaa 3827
Dh 1966 tgcgtatgacttaagaatgacagctgtctactcaaggtatttccaaactatgacacgaa 2025
QY 3828 tggtagatatacagtgataaagtgcgggtctcgggaagggttaaagcgaacgaagcagagct 3887
Dh 2026 tggtagatatacagtgataaagtgcgggtctcgggaagggttaaagcgaacgaagcagagct 2085
QY 3888 gataccccaagagctgagacactgtacatacctgtgcttggaatgagatgtgaataca 3947
Dh 2086 gataccccaagagctgagacactgtacatacctgtgcttggaatgagatgtgaataca 2145
QY 3948 atggaatcccaacaaagacctgaanaattaaagatgatatgttcaacacagaagctgtgtt 4007
Dh 2146 atggaatcccaacaaagacctgaanaattaaagatgatatgttcaacacagaagctgtgtt 2205
QY 4008 cagcagaacatcctcgggaagctcatattgtgctctcgtatgcttcccaatgtccataacc 4067
Dh 2206 cagcagaacatcctcgggaagctcatattgtgctctcgtatgcttcccaatgtccataacc 2265
QY 4068 tgaatcttccacacttggccaacatccacagcacttgaagcggaaattcaacggggcagctt 4127
Dh 2266 tgaatcttccacacttggccaacatccacagcacttgaagcggaaattcaacggggcagctt 2225
QY 4128 catatactgacttggacagctcctcgggaagtgaattatgacacatggaacagctcaacagta 4187
Dh 2226 catatactgacttggacagctcctcgggaagtgaattatgacacatggaacagctcaacagta 2285
QY 4188 tatcatctgaaataagtaacagtatctcttgatcccaagagaacaaatgaaatcctctcca 4247
Dh 2286 tatcatctgaaataagtaacagtatctcttgatcccaagagaacaaatgaaatcctctcca 2445
QY 4248 agtgaatacactgtctccatcccaaggagccaaactcggagaagctctttgtttaa 4307
Dh 2446 agtgaatacactgtctccatcccaaggagccaaactcggagaagctctttgtttaa 2505
QY 4308 accagaanaacttaactttgaaaaatggcacagatcctttcaatgtcatalcaaggctgtga 4367
Dh 2506 accagaanaacttaactttgaaaaatggcacagatcctttcaatgtcatalcaaggctgtga 2565
QY 4368 taaggtcgaatcgaataatccagaataatccaaacttgacagagtatcttgtttatctctcc 4427
Dh 2566 taaggtcgaatcgaataatccagaataatccaaacttgacagagtatcttgtttatctctcc 2625
QY 4428 acagactccgcgaagagacactagctcctgtagaagctgtgcctcctgtccataatcca 4487
Dh 2626 acagactccgcgaagagacactagctcctgtagaagctgtgcctcctgtccataatcca 2665
QY 4488 tatcaacagacacatctcctgcatctcacatttaaaaaatattgtggaagtgtgatagaga 4547
Dh 2686 tatcaacagacacatctcctgcatctcacatttaaaaaatattgtggaagtgtgatagaga 2745
QY 4548 actgacagctgtcaatagcctag 4569
Dh 2746 actgacagctgtcaatagcctag 2767

RESULT 6
AAH33285
ID AAH33285 standard; cDNA; 2867 BP.

XX AAH33285;
AC
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:341.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
RW colorectal carcinoma; chromosome 1; ss.
XX
OS Homo sapiens.
XX
PN MO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137,
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
DR WPI; 2001-235357/24.
DR P-PSDB; AAG73854.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS
XX
Claim 1; Page 2452-2453; 9803pp; English.

CC AAH32943 to AAH37195 and AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX

Sequence 2867 BP; 891 A; 633 C; 643 G; 700 T; 0 other;

	Query Match	54.2%	Score 2477	DB 22	Length 2867	
	Best Local Similarity	91.4%	Pred. No. 0			
	Matches 2766	Conservative	0	Indels 256	Gaps 3	
QY	1544	gagaatcacagagatgtacagcaatgggacatttaagaagttctgtgctacttga	1603			
DB	12	gagaatcacagagagatgtacagcaatgggacatttaagaagttctgtgctacttga	71			
QY	1604	ttcttcacctttagaaggggcccctgagtaattcattctatcagctgaaacaatggt	1663			
DB	72	ttcttcacctttagaaggggcccctgagtaattcattctatcagctgaaacaatggt	131			
QY	1664	atgaagcattgtcgttgaatcagaccccaatgctgcagaaatgtaaaaactctaac	1723			
DB	132	atgaagcattgtcgttgaatcagaccccaatgctgcagaaatgtaaaaactctaac	191			
QY	1724	aataaaggggaggtacacgtcaccagaatgaggaaggagatcgagagaaactcttc	1783			

[illegible]

Db 1113 tcatacagataaacaagctgagctgacagggagacacactgcgcaaaagaattcctcgcagca 1172
 QY 2863 gcttcaaggaggaagctccatctcagcgggcttcgacatctactatgataatgagca 2922
 Db 1173 gcttcaaggaggaagctccatctcagcgggcttcgacatctactatgataatgagca 1220
 QY 2923 cattgctctgttccatgacacacagcagttatgaggagctgcgaacaagaataccaaat 2982
 Db 1221 ----- 1220
 QY 2983 tgggctctctcgcgcgtcagcttagttgtaggaagaatacccaactgatgtctgaa 3042
 Db 1221 -----gtgattaggaagaatacccaactgatgtctgaa 1256
 QY 3043 atgtgctcgtcagcagatgagggaagacacacataaagtgtgtctttaaagagtcaca 3102
 Db 1257 atgtgctcgtcagcagatgagggaagacacacataaagtgtgtctttaaagagtcaca 1316
 QY 3103 caaagtgtgcatcatcatccacacagctcgttggggccctcgtcagctcaagaactagag 3162
 Db 1317 caaagtgtgcatcatcatccacacagctcgttggggccctcgtcagctcaagaactagag 1376
 QY 3163 gagctgtccaaaatgagacagggggtttacagacataatgcttcaagttcaagttcaagaat 3222
 Db 1377 gagctgtccaaaatgagacagggggtttacagacataatgcttcaagttcaagttcaagaat 1436
 QY 3223 ggcctcattgctgtcttctgaggcccttccatcagaagaatgagctcctcagcgtcc 3282
 Db 1437 ggcctcattgctgtcttctgaggcccttccatcagaagaatgagctcctcagcgtcc 1496
 QY 3283 atccagcttgagtaaggaatgaacctccagacagcagctgagtaagtgcagcagtg 3342
 Db 1497 atccagcttgagtaaggaatgaacctccagacagcagctgagtaagtgcagcagtg 1556
 QY 3343 atgtgtgacaagcagctgagggaagaagacaacttgttcttaccactggaacaatacagct 3402
 Db 1557 atgtgtgacaagcagctgagggaagaagacaacttgttcttaccactggaacaatacagct 1616
 QY 3403 ccccaaatcctctctgtgagatccagctgagacagacaggaatgagcttctgtatgtgacaa 3462
 Db 1617 ccccaaatcctctctgtgagatccagctgagacagacaggaatgagcttctgtatgtgacaa 1676
 QY 3463 aacacaaaatgagcctcactccaaatcccaagcattgctaaagtgtgcaacttggaataac 3522
 Db 1677 aacacaaaatgagcctcactccaaatcccaagcattgctaaagtgtgcaacttggaataac 1736
 QY 3523 agtctgacaagaagcttacaacaacttgaacctgaactgacagctccgggtgcctcaatgct 3582
 Db 1737 agtctgacaagaagcttacaacaacttgaacctgacagctccgggtgcctcaatgct 1796
 QY 3583 accctgctccaattacagtgacttccaaaacgaacaagacacacagcaaatcccccagc 3642
 Db 1797 accctgctccaattacagtgacttccaaaacgaacaagacacacagcaaatcccccagc 1856
 QY 3643 cctctgtgagtcttatgcaaatatcgcgaagagcctcccaattccagggccagtgct 3702
 Db 1857 cctctgtgagtcttatgcaaatatcgcgaagagcctcccaattccagggccagtgct 1916
 QY 3703 acaagccttgatgatacagtgtaaatgtaaaacagttccttggagactactgtaaatgga 3762
 Db 1917 acaagccttgatgatacagtgtaaatgtaaaacagttccttggagactactgtaaatgga 1976
 QY 3763 gcaagtgtcgtatgatacagtgtaaatgtaaaacagttccttggagactactgtaaatgga 3822
 Db 1977 gcaagtgtcgtatgatacagtgtaaatgtaaaacagttccttggagactactgtaaatgga 2036
 QY 3823 acaagtgatgatacagtgtaaatgtaaaacagttccttggagactactgtaaatgga 3882
 Db 2037 acaagtgatgatacagtgtaaatgtaaaacagttccttggagactactgtaaatgga 2096
 QY 3883 agagtgatacccccagcagagtgagagcactgtacatactcgtgattgagaatgagaa 3942
 Db 2097 agagtgatacccccagcagagtgagagcactgtacatactcgtgattgagaatgagaa 2156

QY 3943 atacaatggaatccacaagaacctgaaatlaataagatgatagttccaacacaagcagtg 4002
 Db 2157 atacaatggaatccacaagaacctgaaatlaataagatgatagttccaacacaagcagtg 2216
 QY 4003 tgttcagcagaatacctcgggaggtcatttbtgtcttctgatagttccaatgtctcc 4062
 Db 2217 tgttcagcagaatacctcgggaggtcatttbtgtcttctgatagttccaatgtctcc 2276
 QY 4063 atacctatctctccacacctggccaatccacagctggaaggcggaattcaacgggggc 4122
 Db 2277 atacctatctctccacacctggccaatccacagctggaaggcggaattcaacgggggc 2336
 QY 4123 agtctcaatcctgacttggagacagctccttgggagatgatatgacacatggaacagctcac 4182
 Db 2337 agtctcaatcctgacttggagacagctccttgggagatgatatgacacatggaacagctcac 2396
 QY 4183 aagtataatcctggaataagtaacagatcttctgatactcagaagaagttcaatgaaatc 4242
 Db 2397 aagtataatcctggaataagtaacagatcttctgatactcagaagaagttcaatgaaatc 2456
 QY 4243 cttaagtgaaatcactgctcctcatalcccaagaagcacaactcgggagagctcttttg 4302
 Db 2457 cttaagtgaaatcactgctcctcatalcccaagaagcacaactcgggagagctcttttg 2516
 QY 4303 tttaaacagaagaacattacttltgaaatgagcagaacttltcaattgtcattcaagct 4362
 Db 2517 tttaaacagaagaacattacttltgaaatgagcagaacttltcaattgtcattcaagct 2576
 QY 4363 gttgataagtgctgactcgtgaatacagaataatccaaatgtcagagatcttgttatt 4422
 Db 2577 gttgataagtgctgactcgtgaatacagaataatccaaatgtcagagatcttgttatt 2636
 QY 4423 cctccacagactccggcagaagacacactgaactcgtgaataagctgtccttgcctcaat 4482
 Db 2637 cctccacagactccggcagaagacacactgaactcgtgaataagctgtccttgcctcaat 2696
 QY 4483 attcaatcaacagacacattcctcgtgattcaacattttaaataatgagtgagata 4542
 Db 2697 attcaatcaacagacacattcctcgtgattcaacattttaaataatgagtgagata 2756
 QY 4543 ggaagaactgagctgtgaataagcctag 4569
 Db 2757 ggaagaactgagctgtgaataagcctag 2783

RESULT 7
 AAF81927
 ID AAF81927 standard; cDNA; 2745 BP.
 XX
 AC AAF81927;
 XX
 DT 13-JUN-2001 (first entry)
 XX
 DE Human ICAC-1 nucleotide sequence.
 XX
 KW ICAC-1; ICAC-2; asthma; atopic allergy; asthma-related disorder;
 KW interleukin 9 induced calcium activated chloride channel; IL-9;
 KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
 KW anti-inflammatory; immunomodulatory; cystic fibrosis;
 KW inflammatory bowel disease; autoimmune disease; ss.
 XX
 OS Homo sapiens.
 XX
 FT Key Location/Qualifiers
 FT CDS 1..2745
 FT /*tag= a
 FT /product= "ICAC-1"
 FT /note= "IL-9 induced calcium activated chloride channel".
 XX
 PN W09944620-A1.
 PD 10-SEP-1999.

XX 03-MAR-1999; 99MO-US04703.
 PF 03-MAR-1998; 98US-0076815.
 XX (MAGA-) MAGALININ PHARM INC.
 PA Holroyd KJ, Levitt RC, Maloy WL, Lounhed J, McLane M;
 PI Nicotolides NC, Zhou Y, Dong Q;
 XX WPL; 1999-550979/46.
 DR P-PSDB; AAB/4824.
 DR
 PT New nucleic acid encoding calcium activated chloride channel, used to
 PT identify, e.g. specific modulators for treating atopic allergy -
 PS Claim 1; Fig 4B; 75pp; English.
 XX
 CC The present sequence encodes the human interleukin 9 (IL-9) induced
 CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
 CC have anti-allergic, anti-asthmatic, anti-inflammatory and
 CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
 CC used to alleviate asthma (or more generally atopic allergy), while those
 CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
 CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
 CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
 CC (increased levels) or IBD (reduced levels), also for monitoring
 CC treatment of these conditions. The ICACC proteins can be used:
 CC (1) to raise specific antibodies (Ab), useful: (a) as immunoassay
 CC reagents, and (b) as therapeutic (Ab); (11) as fragments or mutants, as
 CC (A) to disrupt bonding between ICACC and its ligands; and (111) to
 CC identify modulators and binding partners. ICACC polynucleotides can be
 CC used to generate transgenic animals or recombinant cells, used to screen
 CC for antagonists, also as a source of therapeutic antisense agents or
 CC diagnostic probes (for quantifying mRNA expression, e.g. for
 CC identification of modulators).
 CC
 SQ Sequence 2745 BP; 833 A; 616 C; 625 G; 671 T; 0 other;
 Query Match 54.1%; Score 2473.6; DB 20; Length 2745;
 Best Local Similarity 96.3%; Pred. NO. 0;
 Matches 2582; Conservative 0; Mismatches 4; Indels 96; Gaps 1;

Db 580 acaaatgtatgaagaagtgctcagggaagcagctgttaaccacaaagaatgacattcaat 639
 Qy 2368 aaagtacaagagctctatgaataaagatgtgtgtttctccaaatcccgcaagcagag 2427
 Db 640 aaagtacaagagctctatgaataaagatgtgtgtttctccaaatcccgcaagcagag 699
 Qy 2428 aagcttctataatgtttgcaacaatgttgaatctatagttgaattcttgcagaacaa 2487
 Db 700 aagcttctataatgtttgcaacaatgttgaatctatagttgaattcttgcagaacaa 759
 Qy 2488 aaccacaacaagaagctctcaacaagaacaaatccaatctcctcgaagcagatg 2547
 Db 760 aaccacaacaagaagctctcaacaagaacaaatccaatctcctcgaagcagatg 819
 Qy 2548 gaagtatccgtgtatctcgaagcttaagaacaacctctatgacaaacagcagca 2607
 Db 820 gaagtatccgtgtatctcgaagcttaagaacaacctctatgacaaacagcagca 879
 Qy 2608 aatccacctctcattgtctcagattgacaaagaattgtgtttgattctctgacaa 2667
 Db 880 aatccacctctcattgtctcagattgacaaagaattgtgtttgattctctgacaa 939
 Qy 2668 tctggaagcattgagcagctgtaacccgctcaatcgactgaatcaagcagcagctttc 2727
 Db 940 tctggaagcattgagcagctgtaacccgctcaatcgactgaatcaagcagcagctttc 999
 Qy 2728 ctgctgcagacagttgagctggtgtcctgtgtttgagatgtgtacatttgacagtgtgc 2787
 Db 1000 ctgctgcagacagttgagctggtgtcctgtgtttgagatgtgtacatttgacagtgtgc 1059
 Qy 2788 catgtacaacaatgtacatacaagataacagttgacagtgacaggaacacatcgccaa 2847
 Db 1060 catgtacaacaatgtacatacaagataacagttgacagtgacaggaacacatcgccaa 1119
 Qy 2848 agatcaccttgacagcagcttccaggaagcagctccatctgcaacggtgttcgtacgtat 2907
 Db 1120 agatcaccttgacagcagcttccaggaagcagctccatctgcaacggtgttcgtacgtat 1179
 Qy 2908 actgatatgtgcaacattgctcgttttccatgacacacagcagttatgggaagtgcga 2967
 Db 1180 act----- 1182
 Qy 2968 caagaanaatccaatctggccctctgctgcagcttagtgattgaggaagaatatcca 3027
 Db 1183 -----gtagatggaagaatatcca 1203
 Qy 3028 actgatatgtgcaacattgctcgttgcagtgacagtgaggaacacacataaagtgtgtgc 3087
 Db 1204 actgatatgtgcaacattgctcgttgcagtgacagtgaggaacacacataaagtgtgtgc 1263
 Qy 3088 tttaacgagtgcaacaagaatgtgtcctcatccacacacagtcgcttggggccctctgca 3147
 Db 1264 tttaacgagtgcaacaagaatgtgtcctcatccacacacagtcgcttggggccctctgca 1323
 Qy 3148 gttcaagaactaggaagagctgtccaaatgacaagaagtttaacaacatatgcttcagat 3207
 Db 1324 gttcaagaactaggaagagctgtccaaatgacaagaagtttaacaacatatgcttcagat 1383
 Qy 3208 caagttcgaacaatlgagctcatltgatttggggcccttccatcaggaatlgagat 3267
 Db 1384 caagttcgaacaatlgagctcatltgatttggggcccttccatcaggaatlgagat 1443
 Qy 3268 gtctctcagcgtctcatccagcttgaagtgtaaggaattaaacctccaaacagcaggtgg 3327
 Db 1444 gtctctcagcgtctcatccagcttgaagtgtaaggaattaaacctccaaacagcaggtgg 1503
 Qy 3328 atgaatgacacagtgatcgttgcagcagcagcttgggaaggaacacttlttctataacc 3387
 Db 1504 atgaatgacacagtgatcgttgcagcagcagcttgggaaggaacacttlttctataacc 1563
 Qy 3388 tggacaatgacagctccccaatccttctctggaatcccaatggaacagaagaagtggtgc 3447

Db 1564 tggacaagcagcctccccaatcctctctgtggtatcccaagtgaacgaagcagcgtgc 1623
 QY 3448 ttgtagtgaacaaaacacccaatggtgctactccaaatcccaagcattgttaagttc 3507
 Db 1624 ttgtagtgaacaaaacacccaatggtgctactccaaatcccaagcattgttaagttc 1683
 QY 3508 ggcacttggaatacagctctgcaagcagctcacaacactgaccctgactgctacgtcc 3567
 Db 1684 ggcacttggaatacagctctgcaagcagctcacaacactgaccctgactgctacgtcc 1743
 QY 3568 cgtgcgtccaaatgctacccctgctcccaattacagtgactcccaaacgaaacagacacc 3627
 Db 1744 cgtgcgtccaaatgctacccctgctcccaattacagtgactcccaaacgaaacagacacc 1803
 QY 3628 agcaaatccccaagcctctgtgtatgttaatgcaaatatgcgaaggagcctcccaatt 3687
 Db 1804 agcaaatccccaagcctctgtgtatgttaatgcaaatatgcgaaggagcctcccaatt 1863
 QY 3688 ctcaaggcccaagctcacaagcctgattgaatcagtgaaatgaaacagttactcttgaa 3747
 Db 1864 ctcaaggcccaagctcacaagcctgattgaatcagtgaaatgaaacagttactcttgaa 1923
 QY 3748 ctactggataatggagcagctgctgactgactaaggatgacggtgctactcaagttat 3807
 Db 1924 ctactggataatggagcagctgctgactgactaaggatgacggtgctactcaagttat 1983
 QY 3808 ttcaacaattatgacacgaatggtatagatagcagtgtaaaagtgcgggctctggagagatt 3867
 Db 1984 ttcaacaattatgacacgaatggtatagatagcagtgtaaaagtgcgggctctggagagatt 2043
 QY 3868 aagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3927
 Db 2044 aagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2103
 QY 3928 attgagaatgatagaataacataggaatcccaagacctgaatataagatgattgtt 3987
 Db 2104 attgagaatgatagaataacataggaatcccaagacctgaatataagatgattgtt 2163
 QY 3988 caacacagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4047
 Db 2164 caacacagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2223
 QY 4048 gtcccaaatgtcccaatcagcagcagcagcagcagcagcagcagcagcagcagcagc 4107
 Db 2224 gtcccaaatgtcccaatcagcagcagcagcagcagcagcagcagcagcagcagcagc 2283
 QY 4108 gaaattcagcgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4167
 Db 2284 gaaattcagcgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2343
 QY 4168 catggagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4227
 Db 2344 catggagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2403
 QY 4228 aagttcaatgaatcctcagcagcagcagcagcagcagcagcagcagcagcagcagc 4287
 Db 2404 aagttcaatgaatcctcagcagcagcagcagcagcagcagcagcagcagcagcagc 2463
 QY 4288 gaggaagcctctgtgttaaacagcagaacatctctgttaaacagcagaagcctcttc 4347
 Db 2464 gaggaagcctctgtgttaaacagcagaacatctctgttaaacagcagaagcctcttc 2523
 QY 4348 atgtgctatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4407
 Db 2524 atgtgctatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2583
 QY 4408 gttatctgtgtttatctcctcagcagcagcagcagcagcagcagcagcagcagcagc 4467
 Db 2584 gttatctgtgtttatctcctcagcagcagcagcagcagcagcagcagcagcagcagc 2643
 QY 4468 gttatctgtgtttatctcctcagcagcagcagcagcagcagcagcagcagcagcagc 4527
 Db 2644 gttatctgtgtttatctcctcagcagcagcagcagcagcagcagcagcagcagcagc 2703

QY 4528 atgtggaatggtgataaggagacagcagcagcagcagcagcagcagcagcagcagcagc 4569
 Db 2704 atgtggaatggtgataaggagacagcagcagcagcagcagcagcagcagcagcagcagc 2745
 RESULT: 8
 AAH46102
 ID AAH46102 standard; DNA: 2742 BP.
 AC AAH46102;
 XX
 DT 11-SEP-2001 (first entry)
 DE Human CLCA1 coding sequence, SEQ ID NO:4.
 XX
 DE Human CLCA1; goblet cell; mouse Gob-5 orthologue; drug screening;
 KW expression inhibition; antisense therapy; gene therapy;
 KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic;
 KW ds.
 XX
 OS Homo sapiens.
 XX
 FH KEY Location/Qualifiers
 FT CDS 1..2742
 FT /tag= a
 FT /partial
 FT /product= "Human CLCA1"
 XX /note= "No stop codon given in the specification"
 PN W0200136530-A1.
 XX
 PD 31-MAY-2001.
 XX
 PF 22-NOV-2000; 2000MO-JP08232.
 XX
 PR 24-NOV-1999; 99JP-0333479.
 PR 27-APR-2000; 2000JP-0127589.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Nakanishi A, Morita S;
 XX
 DR WPI; 2001-355935/37.
 DR P-PSDB; AAB73716.
 XX
 PT New antisense nucleotide, useful for treatment and prevention of
 PT bronchial asthma, and chronic obstructive pulmonary disease -
 XX
 PS Claim 3; Page 82-84; 104pp; Japanese.
 XX
 CC The invention relates to an antisense nucleotide targeted to the mouse
 CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
 CC the CLCA1 gene (coding sequence shown in AAH46102). The invention also
 CC relates to an antibody specific for the Gob-5 protein, medical and
 CC diagnostic compositions containing for the antisense nucleotide, the
 CC antibody, and methods and kits for screening for compounds which inhibit
 CC the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.
 CC The antisense oligonucleotides and antibody are therefore useful for the
 CC treatment and prevention of bronchial asthma and chronic obstructive
 CC pulmonary disease. The present sequence represents the human CLCA1
 CC gene coding sequence.
 XX
 SQ Sequence 2742 BP; 833 A; 616 C; 623 G; 670 T; 0 other;

Query Match 54.0%; Score 2469; DB 22; Length 2742;
 Best Local Similarity 96.2%; Pred. No. 0;
 Matches 2578; Conservative 0; Mismatches 5; Indels 96; Gaps 1;

QY 1888 aaggacatggtgacccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1947
 Db 160 aaggacatggtgacccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 219

OY 1948 ttcaaaaatgttgcattttgatttccttgaacaatgagaaacaaagcgtactatgttga 2007
|||||
Db 220 ttcaaaaatgttgcattttgatttccttgaacaatgagaaacaaagcgtactatgttga 279
OY 2008 ccaaaacttgagaccctcaaaaaatgtctgaatgttctgtgttgcagctactcctcaggt 2067
|||||
Db 280 ccaaaacttgagaccctcaaaaaatgtctgaatgttctgtgttgcagctactcctcaggt 339
OY 2068 aatgtgaacccctactactgagacagatgagcaactgttgaagagaggttgaagaatccac 2127
|||||
Db 340 aatgtgaacccctactactgagacagatgagcaactgttgaagagaggttgaagaatccac 399
OY 2128 ctccctctgtattcatgttcgagaaaagttaactgaatatgtgacccaaggttaaggca 2187
|||||
Db 400 ctccctctgtattcatgttcgagaaaagttaactgaatatgtgacccaaggttaaggca 459
OY 2188 ttgttcacatgagtgaggtcactactacgaltgagagatlattgacgagtaacaataatgtag 2247
|||||
Db 460 ttgttcacatgagtgaggtcactactacgaltgagagatlattgacgagtaacaataatgtag 519
OY 2248 aatctactatcatccaatgtgagaagataataacagtaagatgttccagcggtatactgtgt 2307
|||||
Db 520 aatctactatcatccaatgtgagaagataataacagtaagatgttccagcggtatactgtgt 579
OY 2308 acaaatgtagtaagaagatgtcagagagcgagctgtttacacccaagaatgtcacattcaat 2367
|||||
Db 580 acaaatgtagtaagaagatgtcagagagcgagctgtttacacccaagaatgtcacattcaat 639
OY 2368 aagatlaacagagactctatgaaaaaagatgtgagttgttctccaatcccgccagagagag 2427
|||||
Db 640 aagatlaacagagactctatgaaaaaagatgtgagttgttctccaatcccgccagagagag 699
OY 2428 aaggtcttaataagttttgccaacaatgttgaattcttaagtgtgaatttctgtacagaaacaa 2487
|||||
Db 700 aaggtcttaataagttttgccaacaatgttgaattcttaagtgtgaatttctgtacagaaacaa 759
OY 2488 aaccacaacaagaagagctccacaacaagaatcaaaaatgtcaatctcgcagagccatcgg 2547
|||||
Db 760 aaccacaacaagaagagctccacaacaagaatcaaaaatgtcaatctcgcagagccatcgg 819
OY 2548 gaagtgaatcgttgaattctgaagactttaagaaaacccactctatgaacaacaagccacca 2607
|||||
Db 820 gaagtgaatcgttgaattctgaagactttaagaaaacccactctatgaacaacaagccacca 879
OY 2608 aatcccaacttctatgtgttcagatttgacaagaatgtgtgttttagtctcttgacaaa 2667
|||||
Db 880 aatcccaacttctatgtgttcagatttgacaagaatgtgtgttttagtctcttgacaaa 939
OY 2668 tctggaagcatgagcagctgtgaacgcctcaatcgactgaatcaagcagcgccagcttttc 2727
|||||
Db 940 tctggaagcatgagcagctgtgaacgcctcaatcgactgaatcaagcagcgccagcttttc 999
OY 2728 ctgtgtcagacagatltgagctgtgggttcctgggttgggaatgttgaacatlttgacagtgtgcc 2787
|||||
Db 1000 ctgtgtcagacagatltgagctgtgggttcctgggttgggaatgttgaacatlttgacagtgtgcc 1059
OY 2788 catgtacaataatgaactatatacaataacagtggtcagtgacaggggaacaactgtgccaa 2847
|||||
Db 1060 catgtacaataatgaactatatacaataacagtggtcagtgacaggggaacaactgtgccaa 1119
OY 2848 agattacctgagcagctctcagagggagcgtccatctgcacgaggttcgagctcgagctatt 2907
|||||
Db 1120 agattacctgagcagctctcagagggagcgtccatctgcacgaggttcgagctcgagctatt 1179
OY 2908 actgataltgtgcaacatltgtctgtttccatgacacacagcagttatgtggaggtgca 2967
|||||
Db 1180 act----- 1182
OY 2968 caagaaaaatccaatgtggcctctctgtgcctgagcgttagtgattaggaagaataatcca 3027
|||||
Db 1183 -----gtgattaggaagaataatcca 1203

OY 3028 actgatgatctgaatatgttctgtctgacgagatggggaagaacaactataagtgtgtgc 3087
|||||
Db 1204 actgatgatctgaatatgttctgtctgacgagatggggaagaacaactataagtgtgtgc 1263
OY 3088 tttaacgagatccaacaaagtgtgtccatccacacagttgcgttttgggccccttgcga 3147
|||||
Db 1264 tttaacgagatccaacaaagtgtgtccatccacacagttgcgttttgggccccttgcga 1323
OY 3148 gttcaagaactagagagatgttccaaaatlgacaggaagtttaacagataatgttccagat 3207
|||||
Db 1324 gttcaagaactagagagatgttccaaaatlgacaggaagtttaacagataatgttccagat 1383
OY 3208 caagtctgaacaactatgagctcatgtatgtcttcttgggccccttcaatccaggaatgagct 3267
|||||
Db 1384 caagtctgaacaactatgagctcatgtatgtcttcttgggccccttcaatccaggaatgagct 1443
OY 3268 gtctctcagcgtccatccacagcttgaagatgaaggatlaaacctccagaacagccagttgc 3327
|||||
Db 1444 gtctctcagcgtccatccacagcttgaagatgaaggatlaaacctccagaacagccagttgc 1503
OY 3328 atgaatgacacagatgtgtgtgacagcagcttgggaagaagacattgttcttctaccc 3387
|||||
Db 1504 atgaatgacacagatgtgtgtgacagcagcttgggaagaagacattgttcttctaccc 1563
OY 3388 tggacaatlgcagccctcccaaatccctctctggtgaatcccaatgagcaagaaggtgac 3447
|||||
Db 1564 tggacaatlgcagccctcccaaatccctctctggtgaatcccaatgagcaagaaggtgac 1623
OY 3448 ttgtgtgtgacaaaacacacaaaatgtgcctacccccaatcccgagatgttaaggt 3507
|||||
Db 1624 ttgtgtgtgacaaaacacacaaaatgtgcctacccccaatcccgagatgttaaggt 1683
OY 3508 ggcacttgaatatcaagctgtcgaagcaagctcacaaactgtgacccctgactgctacgtcc 3567
|||||
Db 1664 ggcacttgaatatcaagctgtcgaagcaagctcacaaactgtgacccctgactgctacgtcc 1743
OY 3568 cgtgtccaatgtctaccctgtccccaattacagtgacttccaaaacgaacagagacacc 3627
|||||
Db 1744 cgtgtccaatgtctaccctgtccccaattacagtgacttccaaaacgaacagagacacc 1803
OY 3628 agcaaatccccaagccctctgtgtgttgaatgaacaaatltgcgccaagagccctcccaat 3687
|||||
Db 1804 agcaaatccccaagccctctgtgtgttgaatgaacaaatltgcgccaagagccctcccaat 1863
OY 3688 ctcaagggccagatgttcaacagccctgtatgaatcagtgatgtgaaaaaacagttaccccttgaa 3747
|||||
Db 1864 ctcaagggccagatgttcaacagccctgtatgaatcagtgatgtgaaaaaacagttaccccttgaa 1923
OY 3748 ctactggaataatgagcaggtgtgatagtctactaaggaatgacggtgtactcaaggtat 3807
|||||
Db 1924 ctactggaataatgagcaggtgtgatagtctactaaggaatgacggtgtactcaaggtat 1983
OY 3808 ttcaacaattatgacacgaatgtgtagatagatgttaaaagtgtgaggtctctgggaaggt 3867
|||||
Db 1984 ttcaacaattatgacacgaatgtgtagatagatgttaaaagtgtgaggtctctgggaaggt 2043
OY 3868 aacgagcccaacaggaaggtgtataccccaagcaagatggaagcaactgttaataactgtgcgg 3927
|||||
Db 2044 aacgagcccaacaggaaggtgtataccccaagcaagatggaagcaactgttaataactgtgcgg 2103
OY 3928 attggaatgtgaataacaaatggaatccacaacaaactgtgaataatgaagatgtt 3987
|||||
Db 2104 attggaatgtgaataacaaatggaatccacaacaaactgtgaataatgaagatgtt 2163
OY 3988 caacaacagcaaggtgttctcagcagaacacactcctcgggaaggtcatttgtgtgtctgtat 4047
|||||
Db 2164 caacaacagcaaggtgttctcagcagaacacactcctcgggaaggtcatttgtgtgtctgtat 2223
OY 4048 gtcccaaatgtctccatacctcgatctctcccaactgtgcgaataatccagcactgaagcg 4107
|||||
Db 2224 gtcccaaatgtctccatacctcgatctctcccaactgtgcgaataatccagcactgaagcg 2283
OY 4108 gaattcaagggggagctcacttaactgtgacttgaacagctccctgggaggtatgtac 4167

QY	2670	tggaagacatgagcagctgtgtaacgcgcctccatctgactggaatccaagcagcgcccttccct	2729
Db	829	tggaagacatgagcagctgtgtaacgcgcgcctccatctgactggaatccaagcagcgcccttccct	888
QY	2730	gctgcagacagttgagctctgggctccctggtcttggtatggttgacatttgacagctgctccca	2789
Db	889	gctgcagacagttgagctctgggctccctggtcttggtatggttgacatttgacagctgctccca	948
QY	2790	tgtaacaatggaactccaataagaataaacagctggcagctgacagggacacacccgcgcaaaag	2849
Db	949	tgtaacaatggaactccaataagaataaacagctggcagctgacagggacacacccgcgcaaaag	1008
QY	2850	attaacctgcagcagctctcaggaaggagcgtccactctgcagcagcggtctcgatctgcgattac	2909
Db	1009	attaacctgcagcagctctcaggaaggagcgtccactctgcagcagcggtctcgatctgcgattac	1068
QY	2910	tgatactgtggcaacacttgcctcttctccaatgacacacagcagttatvgggagtgcgcaca	2969
Db	1069	t-----	1069
QY	2970	agaaatcccaattggcgccctctgcgcctgcagcttagtgattgaagaaataatccac	3029
Db	1070	-----gtgattgaagaaataatccac	1092
QY	3030	tgatgtaactctgaaatttgctgctctgcagcgaatgggggaagaacaacatlaaagtggtgctt	3089
Db	1093	tgatgtaactctgaaatttgctgctctgcagcgaatgggggaagaacaacatlaaagtggtgctt	1152
QY	3090	taagaaagtcaaacaaagctggctccatccaataacaaagctgcttggggccctctgcagc	3149
Db	1153	taacgaagtcanaaacaagaatggctgctccatccaataacaaagctgcttggggccctctgcagc	1212
QY	3150	tcaagaactagagagagctgtgcacaaatggcagaaggtttacagaaatctgcttcagatca	3209
Db	1213	tcaagaactagagagagctgtgcacaaatggcagaaggtttacagaaatctgcttcagatca	1272
QY	3210	agttcagaacaatgagcctcaatgtaatgcttcttgggccccttccatcagaagaaatvgagctgt	3269
Db	1273	agttcagaacaatgagcctcaatgtaatgcttcttgggccccttccatcagaagaaatvgagctgt	1332
QY	3270	ctctcagcagctccatccagcttgaaagtgaaaggtataaacccctccgaacagccagcttgat	3329
Db	1333	ctctcagcagctccatccagcttgaaagtgaaaggtataaacccctccgaacagccagcttgat	1392
QY	3330	gaatgtagcacagtgatcgtgtgacagcacgcgtgggaaagacacattgttctctatccactg	3389
Db	1393	gaatgtagcacagtgatcgtgtgacagcacgcgtgggaaagacacattgttctctatccactg	1452
QY	3390	gacaaatgcagcctccccaatctctctctcgtggtaccctcagtgagcagaagatgtgctt	3449
Db	1453	gacaaatgcagcctccccaatctctctctcgtggtaccctcagtgagcagaagatgtgctt	1512
QY	3450	tgtagtgacaacaaacccaataatggccttaactcccaatcccgagcattgtctaaagtgttg	3509
Db	1513	tgtagtgacaacaaacccaataatggccttaactcccaatcccgagcattgtctaaagtgttg	1572
QY	3510	caatttggaaatlacagctctgaaagcagcctccacaacacttgcacccgcagctgtccatcccg	3569
Db	1573	caatttggaaatlacagctctgaaagcagcctccacaacacttgcacccgcagctgtccatcccg	1632
QY	3570	tgctgtccaatgcttaacctgtgcttccatctacatgactgtccaaacagcaagaagacacag	3629
Db	1633	tgctgtccaatgcttaacctgtgcttccatctacatgactgtccaaacagcaagaagacacag	1692
QY	3630	caaatctcccccagcctctgttgattatggaatatctgcgaaggagcctccccaattct	3689
Db	1693	caaatctcccccagcctctgttgattatggaatatctgcgaaggagcctccccaattct	1732
QY	3690	caaggcagatgtcacaagcgccttgaatgaatcagtgaatggaaanaacagttacacttggaaat	3749
Db	1753	caaggcagatgtcacaagcgccttgaatgaatcagtgaatggaaanaacagttacacttggaaat	1812

QY	3750	actggtataatgagcagctgctatgctactataagatlgacggtgtctactcaaggatatt	3809
Db	1813	actggtataatgagcaggtgctgatgtctactaagatlgacggtgtctactcaaggatatt	1872
QY	3810	cacacattatcacccggaatggtgatatacagtgtaaaagctggggcctctggggagagttaa	3869
Db	1873	cacacattatcacccggaatggtgatatacagtgtaaaagctggggcctctggggagagttaa	1932
QY	3870	cgcagccgaagcggagatgtaatacccaagcagatggagacactgtatacataccctgcctgat	3929
Db	1933	cgcagccgaagcggagatggtatgatacccaagcagatggagacactgtatacataccctgcctgat	1992
QY	3930	tgagatgatagaaatacaatactgaaatcccaagaacctgtaataatgaatgatactgtca	3989
Db	1993	tgagatgatagaaatacaatactgaaatcccaagaacctgtaataatgaatgatactgtca	2052
QY	3990	acacagaagaatgtgttcacggagaatcctctcgggaaggtcattttgtgcttcagatgt	4049
Db	2053	acacagaagaatgtgttcacggagaatcctctcgggaaggtcattttgtgcttcagatgt	2112
QY	4050	cccaatgtccccaatactcattctctccaccctggccaatcaacggacctgaagcggga	4109
Db	2113	cccaatgtccccaatactcattctctccaccctggccaatcaacggacctgaagcggga	2172
QY	4110	aattcaacggggcgagctcattcaatctgacatgtagacagctcctcctgggagatgatgacca	4169
Db	2173	aattcaacggggcgagctcattcaatctgacatgtagacagctcctcctgggagatgatgacca	2232
QY	4170	tgggaacgctcacagatattcatttggaataagttacaagtattctttgatacttaagaccaa	4229
Db	2233	tgggaacgctcacagatattcatttggaataagttacaagtattctttgatacttaagaccaa	2292
QY	4230	gttcaatgaatcctctcaagtatgaatactgtcttcacatcccaagaagccaactctga	4289
Db	2293	gttcaatgaatcctctcaagtatgaatactgtcttcacatcccaagaagccaactctga	2352
QY	4290	ggaagctctttgtttaaaccogaanaacattacttttgaaatgtagacagatctttcat	4349
Db	2353	ggaagctctttgtttaaaccogaanaacattacttttgaaatgtagacagatctttcat	2412
QY	4350	tgctattcaagcctgtatataagttcgatctcgaaataccaagaatatccaatctgcagagt	4409
Db	2413	tgctattcaagcctgtgtataagttcgatctcgaaataccaagaatatccaatctgcagagt	2472
QY	4410	atcttgttttatctctccacagactccgcgcagagagacactctgatactgtagaagctctgc	4469
Db	2473	atcttgttttatctctccacagactccgcgcagagagacactctgatactgtagaagctctgc	2532
QY	4470	tccttctcctaataatataatacaacagacacattctctgcatcacattttaaaattat	4529
Db	2533	tccttctcctaataatataatacaacagacacattctctgcatcacattttaaaattat	2592
QY	4530	gtggaagtgatagagaactgcagcctgtataagctgag	4569
Db	2593	gtggaagtgatagagaactgcagcctgtataagctgag	2632
RESULT 10			
AAH46101			
ID	AAH46101 standard; DNA; 2739 BP.		
AC	AAH46101;		
XX			
XX	11-SEP-2001 (first entry)		
DE	Mouse Gob-5 coding sequence, SEQ ID NO:3.		
XX			
KM	Mouse Gob-5; murine; goblet cell; human C1CA1 orthologue; drug screening;		
KM	expression inhibition; antisense therapy; gene therapy; bronchial asthma;		
KM	chronic obstructive pulmonary disease; antiasthmatic; ds.		
XX			
OS	Mus sp.		
XX			

FH	Key	Location/Qualifiers
FT	CDS	1..2739
FT		/tag= a
FT		/partial
FT		/product= "Mouse Gob-5"
FT		/note= "No stop codon given in the specification"
PN	WO200138530-A1.	
XX		
PD	31-MAY-2001.	
PF	22-NOV-2000; 2000MO-JP08232.	
XX		
XX	24-NOV-1999; 99JP-0333479.	
PR	27-APR-2000; 2000JP-0127589.	
XX		
PA	(TAKE) TAKEDA CHEM IND LTD.	
XX		
PI	Nakanishi A, Morita S;	
XX		
DR	WPI; 2001-355935/37.	
DR	P-PsDB; AAB73715.	
XX		
PT	New antisense nucleotide, useful for treatment and prevention of	
XX	bronchial asthma and chronic obstructive pulmonary disease	
XX	Claim 3; Page 80-82; 104pp; Japanese.	
CC	The invention relates to an antisense nucleotide targeted to the mouse	
CC	Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,	
CC	the CLCA1 gene (coding sequence shown in AAH46102). The invention also	
CC	relates to an antibody specific for the Gob-5 protein, medical and	
CC	diagnostic compositions containing the antisense nucleotide or the	
CC	antibody, and methods and kits for screening for compounds which inhibit	
CC	the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.	
CC	The antisense oligonucleotides and antibody are therefore useful for the	
CC	treatment and prevention of bronchial asthma and chronic obstructive	
CC	pulmonary disease. The present sequence represents the mouse Gob-5	
CC	gene coding sequence.	
SQ	Sequence 2739 BP; 784 A; 687 C; 655 G; 613 T; 0 other:	

Query Match	33.2%;	Score 1517.6;	DB 22;	Length 2739;
Best Local Similarity	75.0%;	Pred. No. 0;		
Matches 2017; Conservative	0;	Mismatches 554;	Indels 117;	Gaps 5;

OY	1888	aagsgcatggtgaaccagcgcattctgtatctcgcttgaagcctacaggaaagcgatttat	1947
Db	160	aagsgcatggtgaactcaagcccttcataccggtttgaagcctacaggaaagaattttac	219
OY	1948	ttcaaaaagtgtgcatlittgatcttcctgaacaatggaagcaaacgaagcttacctgtaga	2007
Ddb	220	ttcaaaaagtgtgcatlittgatcttcctgaagagcgtggaaggaacagcctgaatatagag	279
OY	2008	ccaaaaacttgagacctacaanaaaatgctgaatgttctgtgtgctgagtctaactctccaggt	2067
Dd	280	ccaaaacttgagaacctccaanaaagctgagtgcttgttatacaaacagccctctagc	339
OY	2068	aatgatgaaccttaacacgagcagatgggcaactgtgagagaaaggtgaaagatcac	2127
Dd	340	aatgatgagcccttacacgacaatalagagacatgtgagaaaagggagatcagattcac	399
OY	2128	ctcactcctgcatcttattgagcaagaaaaagttagctgaatatagagcacaagagtggca	2187
Dd	400	ctgactcctgcatcttattgagcaagaaagagctgaccagatagggcccaaaagacagacc	459
OY	2188	tttgtcatalagttgggtcatctacagatgggagattttgacagatlacaatatgatgag	2247
Dd	460	tttgtcatalagttgggtcatctacagatgggagagtggttaagaatatacaacaagcagag	519
OY	2248	aaatctactatctcaatggaagaatacaagcaglaagatgttccagcaggtattacttgt	2307

Dh	520	aagttctaacttataccaaagaaaaccccaagcagtgagtgvtltcagcagcaatcaacggt	579
Qy	2308	acaatctgtaaaagaaagctgcagagagagcagctgtatcaaccca---aagatgcacatlc	2364
Dh	580	aaaaatacaagtctcggtcggtgcagagagagcagttgtatcaactaaacgaagaagtgtaac	639
Qy	2365	aataaagtaacagagactctatgnaaaaaagatgtgagttgtcttccaatcccgacagc	2424
Dh	640	gacagagtaacagagactgtataaagaacaattgtgtattgttaacagatccacaaccaaac	699
Qy	2425	gagaaaggttctataaagtgttgcacaacatgttgattctatagttgaattctgtcatga	2484
Dh	700	gagaaaggtcttccatcatatgtttaaaccaataatataatcttctgtgtgtatctctgtaca	759
Qy	2485	aaaaacccaacaagaagaagctccaacaagaacaaatccaanaatgacatctccgaagcca	2544
Dh	760	aaaaatcaacataagaagaccccaaatgacaaacaaacgaatgcaatctccgaagacg	819
Qy	2545	tgggaagtgcacgctgtatcttcagagactttaagaaaaacactctatgacaacagcca	2604
Dh	820	tgggaagatcatccagagatctgagagacttacaagcaaaacactccatgcagccagcca	879
Qy	2605	ccaatctccacactctcatatgcttgcagatcttggaagaagaattgtgtgttaacttacc	2666
Dh	880	ctgtgacccacactcttccacatgcgtgcacaatttgacaagaagatttgtctgaattcttgc	939
Qy	2665	aaatctggaagaagcagctgcagctgtgaacccgctcaatccgactgtaatcaagaacagact	2724
Dh	940	aagttccgagagcatgcgtcgaagaagatgcgtcttcaacccgaatgatatgaagcagcgct	999
Qy	2725	tctcgtcgtcagacagtttgagctgtgggtctcctgggtcttggaatgtgacatttgacagtg	2784
Dh	1000	tctcgtcgtcagagactgtgagacaagagatctcttgggtctggagatgtgacacttgcagtg	1058
Qy	2785	gcccctgtaacaanaatgacatcatatacaataaagaatgagcgtgacaggggacaacatcc	2844
Dh	1060	gctctagttaacaagcgaactcaaacagttaaacaagctgtgtgcagagaaatctgtctgac	1119
Qy	2845	aaaagatataccttgacagcagctctcagagggagagcttccatctgcagcggggtctgcag	2904
Dh	1120	aagcacttaaccacaagatctcgaagggggacatactatgctcttgcgtcccttgagcagca	1179
Qy	2905	tttactgatatgtggtgcaacatltgcctgttttccaatgacacagaagttatcgggagtg	2964
Dh	1180	tttaac-----	1184
Qy	2965	cgacaagaanaatccaatltgggctcctcttgcgtcagcttgatgattagagaanaatat	3024
Dh	1185	-----ggtataaagaagaagatt	1203
Qy	3025	ccaactgtatgtatctgaatatgtgtcgtctbaagatctggtggaagaacaacataagtggtg	3084
Dh	1204	ccaactgtatgtatcttgaatatgtgtcgtctgacccgaatggtggagagacaacacataagcgc	1263
Qy	3085	tgtcttcaacgaagtccaacaagaatgtgtgcataccaacagacgcgtcttggggccctct	3144
Dh	1264	tgtcttgcacgtgtgaaagcagaagcggggtccatcatcatcaagatggtgccttggagccgct	1323
Qy	3145	gcaagctcaagacatagagggggtgtccaanaatgacaggggtttacagacatatgtctca	3204
Dh	1324	gcccgttaaagagcttgggcagcgtgtccaanaatgacaagggggtcctgcagacataactctcg	1383
Qy	3205	gacacagttcagaaacaatgtgcctcatgtgacttcttggggcccttcatcaggaatgtga	3264
Dh	1384	gacacagttcagaaacaatgtgtcttgtatgtcttctgcagacactctctccaggaatgtcg	1443
Qy	3265	gctgtctctcagcgttccatcatcagcgttbgagatgaagaatgaatcaactccagaacagcag	3324
Dh	1444	gcgatctgcacgcactccatccagcgtgtggagagaggtgttaatctccagaataaacaa	1503
Qy	3325	tgtatgtaatgtgcagatgtatcgtgtgacagcacccgtgtggaaagacaacttgttctctatc	3384
Dh	1504	tgtatgtaatgtctcaatgtatcgtgtgacaacgtcgtgtgtgcaagaacacattgttctcttc	1563

OY	3385	aactggacaatgcagcgcctccccaactcctctctcgtgatacccaagtcgagacaagaagaagt	3444
Db	1564	aaccggacaagcagccctccctacatacttatacttggaatccagcgagtcggaaanaaat	1633
OY	3445	gactcttgtaatggaacaanaacacccaataatgagcttacctccaatccagacatctgtaag	3504
Db	1624	gattttactatagaacacaacactaagaatgagtcctactcctcaagtcocagcgacgtctaa	1693
OY	3505	gttcgaccttggaataatcacgtctcgaagcaagctcaaaccttgaacctgaactgtaacg	3564
Db	1664	gttcgaccttggaataatcacgacttcaagcgagctcacaagactcctcaacttgaactgctcc	1743
OY	3565	tcccggtgcctcaatgcttacccgctccctccaattacagttacttccaacaagcaagaagac	3624
Db	1744	tcccggtgcagcaagtgtctaactgcctcctattacagctgaaccccggtgtgaataaagaac	1803
OY	3625	accagcaaatcccccagccctcctgtgagttaatgacaatatctgcgaagagcctcccca	3684
Db	1804	acaaggagaatctcccccagccctctgaacaagttaatgacaagctatgcgaagagcctgcct	1863
OY	3685	attctcagggccaggtgtgcacagcccgattgtaatacgaatgaaatgnaaaaacgttaacctg	3744
Db	1864	attctcagggccccagcgcacaagcctctgatactgcgtgaatgnaaaaacgataacctcg	1923
OY	3745	gaactactggaataatgagacaggtgctctgaatgtaactaaagataagacgtctactcaag	3804
Db	1924	gaattactgtataaagacgagaggtgcgcgtatgcacaaagatatgattggtctactcaag	1983
OY	3805	tatttcaacctatgcacagaatgtaataatagatgtaataaagtgcggtcctctggagga	3864
Db	1964	tttttcaaccttcttgaatgtaaaaatggtatagatcacagcgtttaaaaataatggtcctctggagga	2043
OY	3865	gttaacgacgacacagacgagagatgatacccccagacagtcgagcacttataactcgtgc	3924
Db	2044	gtcaacttcgaacacagacagagacagcaactccgaagaacagagccaatgtaactatgagtc	2103
OY	3925	tgtactgagaatgataatgaataacaaatggaatccacaagaacactgtaaatataaagatgat	3984
Db	2104	tgtactgagaatgataatgaataatgaataacaaacccacgcgtcgtgaactagtct-----at	2157
OY	3985	gttcaacaacaagcagaagtggtttctcagacagaataactcctcgagagctactattgttgctctc	4044
Db	2158	gttcaagaacacagcgcgtggtcttcaagcagaacactcctccaggggagctgtctgtggccaac	2217
OY	4045	gattgc--caaatgctcccatcacctgtacatcttccaccctgcgacaaatacacagcactg	4101
Db	2218	aattgcctccgcagcagatcccatctcctgtgccttctccaccctgtccaatactgactg	2277
OY	4102	aaggcggaattcaagcggggcagcttccatatactgacttggacaagctcctcgggagatgat	4161
Db	2278	aaggcgcgactcacaagggcagaacccctggatcttgaatctgaacggtccctcgggagatgac	2337
OY	4162	tatgacaatggaacaagctcacaagttatcatctggaataagaacagatcttgactctc	4221
Db	2338	tagacacaagcgagagcttccaactacatcatccgaatgagacaagatcgttgactctc	2397
OY	4222	agagacaagtcaatgaatcctcttccaagtgaatactactgctctcatccacaagaagacc	4281
Db	2398	agggacaactacaacactcactccaagtgaaacatacgcgtcttatccacaagaagacc	2457
OY	4282	aactctgagaagaatctctttgtttaaacaagaanaacatnactttgaaatgagcaagat	4341
Db	2458	agctctgagaagaatccttcttgattgtaacgcgggagcaaacactttggaaatgagcaagat	2517
OY	4342	ctttcaatgctcatcaagcgtgtgtgaataagtcgcatgtgaatccgaanaatctcaacat	4401
Db	2518	attctcatgtctatccagcggtgtgataagtcacaatctgaaatccgaanaatctccacaact	2577
OY	4402	gacgaagatactcttgtttatctctccaagaactccgcagagacaacactagtcctgatgaa	4461
Db	2578	gcaacgggtgctgtgttcatcaccgcgtgc-----gacgcgcacatctccgaagaac	2628

Oy	4462	agctgcgcctctgtccatattatcatcaagaagcattccctggacttcaattta	4521
Db	2629	tcaactccctctgtctctcatcattcgaatcaagaagcattccctggacttcaattt	2688
Oy	4522	aaatattatgtgaagtgtgataaggaactgcagctgtcattagacttag	4569
Db	2689	aagataatgtggaagtgtgataaggaactgcagctgtcattagacttag	2736
RESULT 11			
ID	AAH46120	standard; cDNA; 2843 BP.	
AC	AAH46120;		
XX	11-SEP-2001	(first entry)	
DE	Mouse Gob-5 cDNA, SEQ ID NO:22.		
XX			
KW	Mouse Gob-5; murine; goblet cell; human CLCA1 orthologue; drug screening; expression inhibition; antisense therapy; gene therapy; bronchial asthma; chronic obstructive pulmonary disease; antiasthmatic; ss.		
OS	Mus sp.		
PH	Key	Location/Qualifiers	
FT	CDS	15..2756	
FT		/*tag= a	
FT		/product= "Mouse Gob-5"	
PN	WO200138530-A1.		
XX			
PD	31-MAY-2001.		
XX			
PE	22-NOV-2000; 2000WO-JP08232.		
XX			
PR	24-NOV-1999; 99JP-0333479.		
XX	27-APR-2000; 2000JP-0127589.		
PA	(TAKE) TAKEDA CHEM IND LTD.		
XX			
PI	Nakanishi A, Morita S;		
XX			
DR	WPI: 2001-355935/37.		
XX	P-PSDB; AAB73715.		
PT	New antisense nucleotide, useful for treatment and prevention of		
PT	bronchial asthma and chronic obstructive pulmonary disease		
XX			
PS	Example 1; Page 89-91; 104pp; Japanese.		
XX			
CC	The invention relates to an antisense nucleotide targetted to the mouse		
CC	Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,		
CC	the CLCA1 gene (coding sequence shown in AAH46102). The invention also		
CC	relates to an antibody specific for the Gob-5 protein, medical and		
CC	diagnostic compositions containing the antisense nucleotide or the		
CC	antibody, and methods and kits for screening for compounds which inhibit		
CC	the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.		
CC	The antisense oligonucleotides and antibody are therefore useful for the		
CC	treatment and prevention of bronchial asthma and chronic obstructive		
CC	pulmonary disease. The present sequence represents mouse Gob-5 cDNA.		
XX			
SO	Sequence 2843 BP; 815 A; 705 C; 680 G; 643 T; 0 other;		
Query Match	33.2%;	Score 1517.6;	DB 22; Length 2843;
Best Local Similarity	75.0%;	Pred. No. 0;	
Matches 2017; Conservative	0;	Mismatches 554;	Indels 117; Gaps 5
Oy	1888	aagacatggtgaccagcagcatctctgtatctgcttgaagctacaggaagcgatttat	1947
Db	174	aagacatggttacttaccagcagcctctcatcactgttttgaagctacaggaagcgatttac	233

QY	1948	ttnaaaatgltgcatttggatttctctgaaacgtgaaagacaagaagctgcagcatgtgga	2007
Db	234	ttcaaaaatgtgtccattttgatttcccgaaagctgtgaaagcgaaacgtgaaatatacagg	293
QY	2008	ccaaactgtgacctcaaaaatgtcgaatgtctgtgttcgcgagtcctactcccaagt	2067
Db	234	ccaaactgtgaaaccttcaaaaacgtctgattgtctctgtgttcaacaaccagcccttaagc	353
QY	2068	aatgatagcaacctcaactctgcagatggtggcaactgtgaaagaaaggtgaaagatccac	2127
Db	354	aatgatagcccctacacccgaaacatatagagcagtgtggaagaaagggatcagattcac	413
QY	2128	ctcaactctgtatttcatcttcaggaagaaaagttaagctcgaatatgagcacaaagtatggca	2187
Db	414	ctgactctctgactctttagaagaaagaagctgtacacagtatatggtgcacaaagacagacc	473
QY	2188	ttttccatctggtgtggtctctactacgatgtggagatatttgacagtatcaatatatgatag	2247
Db	474	ttttccatctggtgtggtctctactccgactcgcagtgtggaggtttcaatgatacaacagcag	533
QY	2248	aaattctactatccaaatggaagaaatacaagcagtaagatgtctcagcagttacttgt	2307
Db	534	aagtctactatccaaagaaagaaaccccaagcagtgaggtgtctcagcagccattacggt	593
QY	2308	acaaatgtatgtaagaaggtgtctcaggtgagcagctgtttacaacca---aagatgcacattc	2364
Db	594	aaaaatacagttctgtcgtgtgcagaggaagcagattgtatcatcaacgaaagtgttatac	653
QY	2365	aataaagtaaacagaagacctatgaaagaagtgtgagttgtcttccaaatcccgccagagcg	2424
Db	654	gacagagtaacggaacgtgtataagaacaaattgtgtattgtgtccagatccacaaccaaac	713
QY	2425	gagaagagcttctataatgtttgtgcacaacatgtgtatctctatagtttgaattcgttacaag	2484
Db	714	gagaagagcttccatcatcattgtttaaccaaaatataatctctgtgttgaaatctgtacagaa	773
QY	2485	caaaacccaacaagaagagctcccaacaagcaaatccaataatgcatactccgaagaca	2544
Db	774	aaaaatacatcaatcagaagaagccccaatgtgacaaacccaacgatatgtcaatctccgaagcag	833
QY	2545	tgggaagtatcagtgattcttgagagctttaagaagaacacacctatgatacaacaagaca	2604
Db	834	tgggaagtatcagtgaaatccgtgaagacttcaagcaaacacatccacatgacagcccaagca	893
QY	2605	ccaaatccacactctcatatgtctgcagcagatgtgacaagaagattgtgtttagtcccttac	2664
Db	894	cttgaccacacactctcatcagctgcgcgaaatgtgacaagaagattgtgtcttagtcttgat	953
QY	2665	aaatctggaagcatgbcgactgtgtaaccgcctccatcagactgaatcaagcagccagctt	2724
Db	954	aagctccggagcatgcatggaacgatatgtctgtcttaaacccgaatatgaatcaagccagctgt	1013
QY	2725	ttctctgtcgcgaagaagtgtgagctgtggggtctcctggtttgggaatgtgaactttgaacgtct	2784
Db	1014	ttctctgtcgcgaagaagtgtgagcagagatccctgtgtctgggaatgtgacactttgaacgtct	1073
QY	2785	gcccattgtacaaaatgaaactatagcatataaacaagtgtgcagtgaaagggacacactcc	2844
Db	1074	gctcatgtacaagaagcgaaccccaaaacagttaaacagtggtgtcgcagacaagatctgtcatc	1133
QY	2845	aaaaagattacctgcagcagcttcaaggaaggaagtcacatctgcagcgggtctcgatcgca	2904
Db	1134	aagaactatccacaatgatactcgaaggaagacatatagctctcgtcgtctcgacagca	1193
QY	2905	tttactgatatgtggaacaactttgcctgttttccatagtaacacagattatggggagtg	2964
Db	1194	ttttac-----	1198
QY	2965	gcacagaagaataatccaatgtgscctctctgcgcctgcagcttagtgatttagaagaataat	3024
Db	1199	-----agtgataaagaagaagtat	1217
QY	3025	ccaactgtatgatactgaaattgtctgtccgcagcagatgtggaagaaacaaactataagtgg	3084

Db	1218	ccaacgtatgatactcgaaatctgtcgtcgtacccgaatcgaggagaaacaacccattagacgc	1277
QY	3085	tgctttacagcggtgcacaacaagaatgtgtgcatactatccacaagtcgtcttgggcctct	3144
Db	1278	tgctttgacgcggtggaagcagagcgaggccatcatccataacagctgtgcccctggagcgcgt	1337
QY	3145	gacagctcaagaactagagsgcgtgccaaaatgacagsggttacaagacabatgtctca	3204
Db	1338	gcgctcaaaagsgcttgagcagctgtcccaaaatgacagsggcgtccgaacatactcttcg	1397
QY	3205	gataaagttcaagaaacaatbtgcccattgtatgtcttggggcccttcaatcngnaatgtga	3264
Db	1398	gataaagttcaagaaacaatbtgctctgtttatgtcttccgcagctctctccacgaaatgcg	1457
QY	3265	gctgtccctcaagcgtccatccacagcttgagagatgaagaaatlaacccctcagaaacgcgc	3324
Db	1458	gcgatcgctccagcactccatccacgcgtcgagagcagggaggttaactccagataaccaa	1517
QY	3325	tgagaatgtgcacagtgtactcgttggaacagacacgctgggaaagagacacttghttctatc	3384
Db	1518	tgagatgaatgtgcgtcagtgatctcgtggaacagctcgttggaagagacacttghttctatc	1577
QY	3385	accctggacaatgtcagcgtctccccaatccctctctctgtgataccagtgagacagaagaagt	3444
Db	1578	accctggacaagcgtcctctccatacatatctatcttggaatcccaagcgagtggaacaatat	1637
QY	3445	gacctttagtgagacaanaacccaataatgtgcctacatcccaatcccgagcatgtgaag	3504
Db	1638	gctttatatactagacacaacaacctaaatggtgtgcctacatcccaatcccgagcaggtctag	1697
QY	3505	gttgcgaccttggaataatacagctctgtcgaagaagctcacaaccttgaccctgtactgtcag	3564
Db	1698	gttgcgttttggaataatacagcatlccagacgagctccacagactccacatgtactgtccac	1757
QY	3565	tcccggtcgtlccaatgtctacccgtcgtcccaattacagtgacttccaanaagacaagagac	3624
Db	1758	tcccggtcagcaagtgctactacgtcctccatctacagtgagcccggtgtggaataagac	1817
QY	3625	acccgacaattccccacgcctctcgtgtagttatgtacaatattccgacaagcgccctccca	3684
Db	1818	acagaggaattccccacgcctgtgaacaagtgtatgtcagaagcatctgcgcaagggcctctgcct	1877
QY	3685	attctcagggccagtgctacacagccctgtatgtacagtgaaatgynaanaaacagttactctg	3744
Db	1878	attctcagggccagcgctacacagccttgatgtgatactgtgaatgynaanaaacagttactctg	1937
QY	3745	gaactactgtgaataatgtgagcaggtgtctgtatgtactaagaatgtacgcgtgtctactcaagg	3804
Db	1938	gaactactgtgaataatgtgagcagcgtgtcgtatgtccacaagaataatgtatgtctactcaagg	1997
QY	3805	tatttccaacttaatgacacgaaatgtgtaataacagtgtaataaagtgccgggcctctggagga	3864
Db	1998	tatttccaacttaatgacacgaaatgtgtagataacagcgttaaaatatgtggcctctggagga	2057
QY	3865	gttaacgcagccagacgagagatgtataccccaagcagagtgtgacactgttaataactgtgc	3924
Db	2058	gtcaacttcagacagacagagagatgtacacttcgaagaacagagcgatgttaataatgtgc	2117
QY	3925	tgagtatgaagaatgtgaatatatacgaatgtgaatcccaagaagcgtgaatatataagaatgat	3984
Db	2118	tgagtatgaagaatgtgtgaagaatgaagaatgtgaacccaacagcgtctctgtgaatactgat	2171
QY	3985	gttaacaacaagaagaatgtgtttcttcagacagaatatctcgtggaggtctatgttgtctct	4044
Db	2172	gttcaagaacaagaagcgtgtgtcttcagacagaatatcttccagggaggtctgttgtgtccac	2231
QY	4045	gatactc---caaatgctccacataactgtctctccacacgtgcgaatatcacagcactg	4101
Db	2232	aatgtcccgagcagctcccaatctctgtacactcttccacacgtgtccaatactcagcactg	2291
QY	4102	aagcggaatattcaaggggcagcttcatataatctgtacttggacagcctccgtggatgat	4161

OY	2545	tgaaagatgacccgttgctcttgagagactttaaagaacaaacaccccttatgtgacaacagca	2604
Db	827	tggaagaagcacaaccgaagcttgcagagatcttaagcaaacacaccccttcagtcagccagca	886
OY	2605	ccaatccaccccttcatactgtctgcagatctgacaaagaattgtgttaagtccttgac	2664
Db	887	cctgcaaccacacttctccactgttcgccaatttgacaaagaattgtgttgcttaagtc	946
OY	2665	aaatctggaagcatgctgcgactgtgtaaacgcgcctcaatcgcactgaatcaagcagccacgtt	2724
Db	947	aagtcgcggagcatgctgaaacgatactgctcttaacccggaatgaatcaagcaagccgctct	1006
OY	2725	ttccctgcgcgcgaagaagtctgagcgtgagggtcccttggttggtgaattgtgacatttgacagtgct	2784
Db	1007	ttccctgcgcgcgagactgtgagagcagagatccctggtgttcggatgtgtgacctcttgacagtgct	1066
OY	2785	gcccattgbcacaaatgacacccatcacagataaacaagtgcagctgcagcagggacacacatccgc	2844
Db	1067	gctcatgtacacaagcgcgaatccaaacagttaaacagctgctgcagcagagatctgtgtac	1126
OY	2845	aaaagattacctgcagcagcattcagagagagcgtcccatctgcagcgcgctcttgatcgcga	2904
Db	1127	aagacattacccacacagatattcgcagagagagacatactatactgtcctgcgccttcgcgaacga	1186
OY	2905	tttactgatatgtgagcaaatcttgccgtgttcctcatgacacacaagagttatggtgagtg	2964
Db	1187	tttaac-----	1191
OY	2965	cgaagaagaacaaatcccaaatctgagcctctctgcccctgcagactgattgatagagaacaaat	3024
Db	1192	-----agtataagaagaagat	1210
OY	3025	ccaaactgatatgatactgaaattgtctgtctgcgaagatctgaggaagacaacacataaagtgtg	3084
Db	1211	ccaactgatatgatactgaaattgtctgtctgcgaagatctgaggaagacaacacccattagcagc	1270
OY	3085	tgctttaagcagagtcacaaacaaagtgtgtgcatactcaacacagctcgtcttggtggccct	3144
Db	1271	tgcttgcgcctgtgtgaagcgcgaagcggggtccatcatccatacacaagctgtgcccctggagccgct	1330
OY	3145	gcagctcacaagacacagagagcgtctcccaaatctacagagaggtttacagacatatgtctca	3204
Db	1331	gcgcgttaagagcctctgagcgcgctgtcccaaatctgacagaggggcctgcagacataactcttcg	1390
OY	3205	gatacagttcgaacaaatctgcccatactgatatgctcttttggtggcccttcatacaggaatgga	3264
Db	1391	gatacagttcgaacaaatctgatatgctctttcgcagcactcctccctcaggaatatctg	1450
OY	3265	gctgtctctcaagcgtcccatccagcgtttggaagatgaagatataccctccaggaacgcgag	3324
Db	1451	gcgaatgcctcagcctccatccatccagcttgcgaagagcaggttaattccagaataacaa	1510
OY	3325	tggaatgaaatgcacaagctgctgcagcagcaccgttggaagagacacattgttcttatc	3384
Db	1511	tggaatgaaatgcctcagctgctgcagcagcctggtgagcaagcaccctgttcttatc	1570
OY	3385	aacttgacaatctgacgctccccaatcctctctcttggaatccagctgagacagaagaagt	3444
Db	1571	aacttgacaacagcctcctcctcaacaatatattatcttggaatccagcggagtggaacaaat	1630
OY	3445	ggcttttgatgagtgacacaaacacacaaatctgactactcccaatcccgagcttgtcag	3504
Db	1631	ggttttatactatagacacaacacacaaatgagtggtctactcccaatcccgagcccggtctaa	1690
OY	3505	gttggaactttggaatacagctctgcgaagcaggttcaacacacttgacccttgactgacg	3564
Db	1691	gttggaacttttggaatacagcatccaatccaagcaggtccacagactcccaacttgactgtccac	1750
OY	3565	tcocgtgtgtccaaatgtctacccctgcctcccatatagctgaattccaaacagaagacagac	3624
Db	1751	tcocgtgtgaggaagtgtctacacgtgcctccctattaaagctcaacccggtgtagtgaatagac	1810
OY	3625	accagcaaatctccagccctctgtagtattatgcaaatattctgcgaagagagctcccca	3684

[illegible]

Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
pharmaceutical; receptor immunoadhesin; gene mapping; ss.

Homo sapiens.

MO9963088-A2.

09-DEC-1999.

02-JUN-1999; 99WO-US12252.

02-JUN-1998; 98US-0087607.

02-JUN-1998; 98US-0087609.

02-JUN-1998; 98US-0087759.

03-JUN-1998; 98US-0087827.

04-JUN-1998; 98US-0088021.

04-JUN-1998; 98US-0088025.

04-JUN-1998; 98US-0088028.

04-JUN-1998; 98US-0088029.

04-JUN-1998; 98US-0088030.

04-JUN-1998; 98US-0088033.

04-JUN-1998; 98US-0088326.

05-JUN-1998; 98US-0088167.

05-JUN-1998; 98US-0088202.

05-JUN-1998; 98US-0088212.

05-JUN-1998; 98US-0088217.

09-JUN-1998; 98US-0088655.

10-JUN-1998; 98US-0088722.

10-JUN-1998; 98US-0088730.

10-JUN-1998; 98US-0088734.

10-JUN-1998; 98US-0088738.

10-JUN-1998; 98US-0088740.

10-JUN-1998; 98US-0088742.

10-JUN-1998; 98US-0088810.

10-JUN-1998; 98US-0088811.

10-JUN-1998; 98US-0088824.

10-JUN-1998; 98US-0088825.

10-JUN-1998; 98US-0088826.

11-JUN-1998; 98US-0088838.

11-JUN-1998; 98US-0088861.

11-JUN-1998; 98US-0088863.

11-JUN-1998; 98US-0088876.

12-JUN-1998; 98US-0089090.

12-JUN-1998; 98US-0089105.

16-JUN-1998; 98US-0089440.

16-JUN-1998; 98US-0089512.

17-JUN-1998; 98US-0089514.

17-JUN-1998; 98US-0089532.

17-JUN-1998; 98US-0089538.

17-JUN-1998; 98US-0089598.

17-JUN-1998; 98US-0089599.

17-JUN-1998; 98US-0089600.

17-JUN-1998; 98US-0089653.

18-JUN-1998; 98US-0089801.

18-JUN-1998; 98US-0089907.

19-JUN-1998; 98US-0089947.

19-JUN-1998; 98US-0089948.

19-JUN-1998; 98US-0089948.

22-JUN-1998; 98US-0090246.

22-JUN-1998; 98US-0090252.

22-JUN-1998; 98US-0090254.

23-JUN-1998; 98US-0090343.

23-JUN-1998; 98US-0090355.

24-JUN-1998; 98US-0090429.

24-JUN-1998; 98US-0090431.

24-JUN-1998; 98US-0090444.

24-JUN-1998; 98US-0090445.

24-JUN-1998; 98US-0090461.

24-JUN-1998; 98US-0090472.

24-JUN-1998; 98US-0090535.

24-JUN-1998; 98US-0090538.

24-JUN-1998; 98US-0090540.

24-JUN-1998; 98US-0090557.

25-JUN-1998; 98US-0090676.

25-JUN-1998; 98US-0090678.

25-JUN-1998; 98US-0090688.

25-JUN-1998; 98US-0090690.

25-JUN-1998; 98US-0090691.

25-JUN-1998; 98US-0090694.

25-JUN-1998; 98US-0090695.

26-JUN-1998; 98US-0090696.

26-JUN-1998; 98US-0090862.

01-JUL-1998; 98US-0091358.

01-JUL-1998; 98US-0091360.

01-JUL-1998; 98US-0091544.

02-JUL-1998; 98US-0091478.

02-JUL-1998; 98US-0091486.

02-JUL-1998; 98US-0091519.

02-JUL-1998; 98US-0091626.

02-JUL-1998; 98US-0091628.

02-JUL-1998; 98US-0091633.

02-JUL-1998; 98US-0091646.

07-JUL-1998; 98US-0091978.

09-JUL-1998; 98US-0091982.

10-JUL-1998; 98US-0092182.

20-JUL-1998; 98US-0092472.

30-JUL-1998; 98US-0093339.

04-AUG-1998; 98US-0094651.

04-AUG-1998; 98US-0095282.

04-AUG-1998; 98US-0095285.

04-AUG-1998; 98US-0095301.

04-AUG-1998; 98US-0095302.

04-AUG-1998; 98US-0095318.

04-AUG-1998; 98US-0095321.

10-AUG-1998; 98US-0095325.

10-AUG-1998; 98US-0095916.

10-AUG-1998; 98US-0095929.

11-AUG-1998; 98US-0096012.

11-AUG-1998; 98US-0096143.

11-AUG-1998; 98US-0096146.

12-AUG-1998; 98US-0096329.

17-AUG-1998; 98US-0096757.

17-AUG-1998; 98US-0096766.

17-AUG-1998; 98US-0096768.

17-AUG-1998; 98US-0096773.

17-AUG-1998; 98US-0096791.

17-AUG-1998; 98US-0096867.

17-AUG-1998; 98US-0096891.

17-AUG-1998; 98US-0096894.

17-AUG-1998; 98US-0096895.

17-AUG-1998; 98US-0096897.

18-AUG-1998; 98US-0096949.

18-AUG-1998; 98US-0096950.

18-AUG-1998; 98US-0096959.

18-AUG-1998; 98US-0096960.

18-AUG-1998; 98US-0097022.

19-AUG-1998; 98US-0097141.

20-AUG-1998; 98US-0097218.

24-AUG-1998; 98US-0097661.

26-AUG-1998; 98US-0097951.

26-AUG-1998; 98US-0097952.

26-AUG-1998; 98US-0097954.

26-AUG-1998; 98US-0097955.

26-AUG-1998; 98US-0097971.

26-AUG-1998; 98US-0097974.

26-AUG-1998; 98US-0097978.

26-AUG-1998; 98US-0097979.

26-AUG-1998; 98US-0097986.

31-AUG-1998; 98US-0098014.

31-AUG-1998; 98US-0098525.

PR 16-SEP-1998: 980S-0100634.
PR 12-JAN-1999: 990S-0115565.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Matanabe CK;
PI Wood WI, Yuan J;
XX
DR MPI: 2000-072883/06.
DR P-PSDB: AAY66749.
XX
PT Membrane-bound proteins and related nucleotide sequences -
XX
PS Claim 2: Fig 273: 822bp: English.
XX
CC The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block
CC receptor-ligand interactions. The membrane-bound proteins can also be
CC employed for screening of potential peptide or small molecule inhibitors
CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.
XX
SQ Sequence 3265 BP: 1159 A: 596 C: 632 G: 878 T: 0 other:

Query Match 24.7%; Score 1129.8; DB 21; Length 3265;
Best Local Similarity 67.1%; Pred. No. 2e-300;
Matches 1749; Conservative 0; Mismatches 747; Indels 111; Gaps 6;

OY 1889 aggcacatggtgacccagcattctctgtatctgttgagtcacaggaagcgattttatt 1948
DB 182 aggatatgtagtaccagctctcaagctgttgtagcagcagaagaaagattttt 241
OY 1949 tcgaaatgtgcatgttgatcttcctggaacatggaagaaagcgctgatagttgagc 2008
DB 242 tcaaaatgtatcatatataatcctcctgagatcggaagaaacccctcagtaacaaagcg 301
OY 2009 caaaacttgagcactcaaaaatgctatgtctcgtgtgtagctacacccctccagta 2068
DB 302 caaaacttgagcactcaaaaatgctatgtctcgtgtgtagctacacccctccagta 361
OY 2069 atgactgaacctcactcactgagcagatgagcagactgtaggagaagggctgaagatccacc 2128
DB 362 gagaatgaaccatcacacacacagcttcacagatgtaggagaagggatatacttacc 421
OY 2129 tcaactctgttcatatgtagcaggaagaaagttagtgaattgagcccaaggttagggcat 2188
DB 422 tccacctgacctcttactcttggaaaaaacaatgtaattgtagccacccagcaactgt 481
OY 2189 ttgtcactggtggtgcctacatcgatgtaggagatatttgacaggtacaataatgataga 2248
DB 482 ttgtcactggtggtgcctacatcgatgtaggagatatttgagtagacaaatgaagatcagc 541
OY 2249 aattctactatcccaatggaagatac--aagcagtaagatgttccagcagttacttg 2305
DB 542 ctcttctacccgtgctgaagtaacaaaaaatcgaaagcaagtgcttcgcagagtattctctg 601
OY 2306 gtccaatgtagtaagaagtgtcagggagcagctgttaccacaaagtgcacattca 2365
DB 602 gtgaatataagattataaagtgtaagaagcgagctgtcttagtaggcatacgagaattg 661
OY 2366 ataaagtaacaggactctatgaaagaatgtagttgtttcccaatcccgccagacg 2425
DB 662 attctacaacaaaactgtatggaagaagatgtcaattcttctctgataaagttacaacag 721

OY 2426 agaaagcttctataatggttgcacaaatggttattctatagttgattctgttacaagac 2485
DB 722 aaaaagctcacaataatgattatbcaaaagtattgattctgttctgttgaatttgaaga 781
OY 2486 aaaaacccaacaaagaagcctccaaagaagaaatcaaaaatgcaatctccgaagcacat 2545
DB 782 aaacccaataacagaagcctccaaagcctcaacataaagaatgcaatctttagaagatcat 841
OY 2546 gggagatgtagcgttgattctgaagactttaagaaaacacactctcatagacacagccac 2605
DB 842 gggaggtgtagcaattctcgagattttaaaacaccataccatcgtagacacaccc 901
OY 2606 caaatcccaactctcaattgtctgacagattggaacaaagaattgtgttttagtcttgaca 2665
DB 902 ctccacctgtcttcatattgtctgaatcagtaagaagaattgtgtcttagtcttgata 961
OY 2666 aactctgaagcagtgagcagctgtaaccgctcctacatcgactgaatcaagcagccactt 2725
DB 962 agtcggaagcagtgagggggtgaagagccgctcaatcgatgaatcaagcagcaaacatt 1021
OY 2726 tctctgtgcagacagttgagctgggttcctgggttggagatggtgacattgacagtgctg 2785
DB 1022 tctctgtgcagacagttgaaatggaatgcctcgtgggtgagatggttcaacttgatagctg 1081
OY 2786 cccattacaaaatgtaacatcatcaataaagctgagcagtgagcagggacacactgccaa 2845
DB 1082 cccattgtgaataagctaatcacaataaagaacagtgatgaaagaacacacatctgag 1141
OY 2846 aagatctactcagcagcagctctcagagggagcgtccatctgcagcagcgtctgagcagc 2905
DB 1142 caggtattaccatacatctctcgtggaaggaacttccatctgctctggaatataatgcat 1201
OY 2906 ttactgatalgtagcaacatttgcctgttttccatgacacacagcagatagggagctg 2965
DB 1202 ttcaagtgatgtagagc----- 1219
OY 2966 gacaagaanaatccaattggcctctcgtcgtcgcagctgagttagtagaagaatatac 3025
DB 1220 -----tacttccc 1228
OY 3026 caactgtagctatgtaaatgtgtgtctgtacgagatgggagaacacacataatgtagt 3085
DB 1229 aactctgtagtaccgaaagtctgtgtgtaactgtaggaggaataacacgtgcaagttct 1288
OY 3086 gcttcaagagtgcaaaaagctggtgcaatcaccacacagctgcttggggccctctg 3145
DB 1289 gttatgtagagtgaaacaagaatggggccattgttcatattgttcttgggaagcgtg 1348
OY 3146 cagctcaagaactagagagagctgtccaataatgacagagaggtttaacagacatagcttcag 3205
DB 1349 ctgattgaagcagtaataatagatagagaatgaagaatacagaaggaagctatattatgtctcag 1408
OY 3206 atcaagttcagaacaatgagcctcatgtatgcttcttggggccttcatcaggaatggag 3265
DB 1409 atgaagctcagaacaatgagcctcatgtatgcttcttggggccttcatcagaagaaatcag 1468
OY 3266 ctgtctctcagcgtccatccatccagcttgagagtaaggatataacccctcagaagccagt 3325
DB 1469 atctctccagaagtgctccatccagctcgaagaagtaaggatataacacatgataatgctc 1528
OY 3326 ggaatgaatgcaacagtgatcgttgagacagcagctgggaagaagacacttgttctcatca 3385
DB 1529 ggaatgaatgcaacagtgatcgttgagacagcagctgggaagaagacacttgttctcatca 1588
OY 3386 cctggaacaatgcaacagctcccaatctctctcgtggatcccaatggaagagaagctg 3445
DB 1589 catggaacaatgcaacagctcccaatctctctcgtggatcccaatggaagagaatgga 1648
OY 3446 gctttagtgagcaaaaacacaaaatggtcctaccatccaaatcccgagatgtgtaag 3505
DB 1649 attcaacagtgagtgcaacttccaaaatggtcctatctcagtatctccaagaacgtcgaag 1708
OY 3506 ttggcacttggaataacagctgc-----aagcaagctcaacaaccttgacccctgactg 3559

Dh	1709	ttggcacttgggcatcaatctcaatcttcaagcccaagcgaaacccagaaacatlaacttcaag	1768
QY	3560	tcaagtccttggtgtcccaatgtctacccctgtccctccaattacagtgacttccaacagaca	3619
Dh	1769	taactctcgagcagcaaatcttctgtgtccctcaatcacagtgatgtctaataagtata	1828
QY	3620	aggaacacagcaaatccccagccctctgtgtattatgtcaaatcttcgcgaagagcct	3679
Dh	1829	agggcgttaaacagttccccagcccccaatgatgttttlaacgagaaattctlaaagagatgt	1888
QY	3680	ccccaatcttcagggccaggtgtacagcccttgattgaatcagatgtgatgtgaaacagtta	3739
Dh	1889	tacctgtctcttgagcccaatgtgtactgccttcaattgaatcacagaatgtgacatacgaag	1948
QY	3740	ccttggaactactgtgaabaatgtgacaggtgtgcgtgtactaagatgtacgtgttact	3799
Dh	1949	ttttggaacttttgataatgtgtgcagggcgtgatcttcttcaagaatgtatgtact	2008
QY	3800	caagatgtttccaacttatacgaatgtgagtaagtaagtaagttgaagttgggtctgtg	3859
Dh	2009	ccaagatatttcaacatatacagaagaatgtgagataatagctttaaagttcgggtcatgt	2068
QY	3860	gaggaggttaacgcagcccaagacgtagatctaccccaagagatgtgagcactgtacatac	3919
Dh	2069	gaggaggaacacactgcgcaggtctaataattacggtccctccactgaatagcgcgftacac	2128
QY	3920	ctgtgctgtatttggaatgtatgaatacgaatgtgaatcccaacgaagccttgaattataag	3979
Dh	2129	caagcgttggtaagtgaaacggggaatattgaagcaaacccgcgaagacctaataatgat---	2185
QY	3980	atgagtgttcaacaacaagcaagttgttttcaagacagacatccctcgggaagccattgttg	4039
Dh	2186	aggtactcaagacacacctgtggaggtattcagcgcgaacgacatccggaggttgatttgtg	2245
QY	4040	cttctgtatgtcccaaatgtctccatactctgtactcttcccactgtgccaatccgcgacc	4099
Dh	2246	tattacaagctcccaagccttcccttccgtgcaccaatacccaacgaagtaataatcaacagc	2305
QY	4100	tgaagcgaggaaattccacggggcgagctcatattcatctgactgtgacagcttcctgggatg	4159
Dh	2306	ttgatgtccacagtcatgtagag---ataagattatcttcatatgtgcagccacaggagata	2362
QY	4160	attatgaccaatgtgaacagctccaaagttatcatattcogaataagtaacaagtaattcttgc	4219
Dh	2363	atttgtgtgtgaaagaagttcaacggttatcatataagaataagtagtcaagtattctgtac	2422
QY	4220	tcgaagacaaagtccaatgaatctctcttcaagtgaaatactactgtctctatcccaagaag	4279
Dh	2423	taagaagcagttcttgatgtgccttccaagtaatactactactgtgtcatccaagaagag	2482
QY	4280	ccaactcttgaggagctttttttttaaacccgaaacaacttactcttgaataatgtgcag	4339
Dh	2483	ccaactccaagaagcgttgtgaatttaaacccgaataatcttcagaagaataatgtcaacc	2542
QY	4340	atctttcatgtctatcgaagctgtgtgaataaggtcgaatcgtgaatccgaataatccaaca	4399
Dh	2543	acattatttgcatttaagaagatagataaagaacatltgtgcatacaaaatattccaaca	2602
QY	4400	ttgacagagatcttgtttattctccccaagacatccgcgcagaga---cacctgtctgt	4456
Dh	2603	tttgacaagaacttgtttattatccctccaagccaatcccgatgtgacattgactctacacta	2662
QY	4457	atgaaacgtctgtcctgttccataa	4483
Dh	2663	ctctactactactcctacactcgtgata	2689

RESULT	14
AAAF92092	
ID	AAAF92092 standard; cDNA; 3265 bp
XX	
AC	AAAF92092;

DT	15-MAY-2001	(first entry)	
XX			
DE	Human PRO1124 cDNA.		
XX			
KW	Human; PRO protein; mapping; ss.		
XX			
OS	Homo sapiens.		
PN	WO200116318-A2.		
XX			
PD	08-MAR-2001.		
XX			
PE	24-AUG-2000; 2000WO-US23328.		
XX			
PR	01-SEP-1999; 99WO-US20111.		
PR	15-SEP-1999; 99WO-US21090.		
PR	07-DEC-1999; 99US-0169495.		
PR	09-DEC-1999; 99US-0170262.		
PR	11-JAN-2000; 2000US-0175481.		
PR	18-FEB-2000; 2000WO-US04341.		
PR	18-FEB-2000; 2000WO-US04342.		
PR	22-FEB-2000; 2000WO-US04414.		
PR	01-MAR-2000; 2000WO-US05601.		
PR	03-MAR-2000; 2000US-0187202.		
PR	25-APR-2000; 2000US-0199397.		
PR	22-MAY-2000; 2000WO-US14042.		
PR	05-JUN-2000; 2000US-0209832.		
XX			
PA	(GENE) GENENTECH INC.		
XX			
PI	Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;		
PI	Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;		
XX			
DR	WPI: 2001-183260/18.		
DR	P-PSDB: AAB87560.		
XX			
PT	Eighty four nucleic acids encoding PRO polypeptides, useful in		
PT	molecular biology, including use as hybridization probes, and in		
PT	chromosome and gene mapping.		
XX			
PS	Claim 2; Fig 69; 278pp; English.		
XX			
CC	The present sequence is the coding sequence for a human PRO polypeptide		
CC	(secreted and transmembrane). The PRO protein, and PRO agonists, PRO		
CC	antagonists or anti-PRO antibodies are useful for preparation of a		
CC	medicament useful in the treatment of a condition which is responsive to		
CC	the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO		
CC	protein may also be employed as molecular weight markers for protein		
CC	electrophoresis. The PRO coding sequence has applications in molecular		
CC	biology, including use as hybridisation probes, and in chromosome and		
CC	gene mapping.		
XX			
SQ	Sequence 3265 BP; 1159 A; 596 C; 632 G; 878 T; 0 other;		
XX			
Query Match	24.7%; Score 1129.8; DB 22; Length 3265;		
Best Local Similarity	67.1%; Pred. No.2e-300;		
Matches 1749; Conservative	0; Mismatches 747; Indels .111; Gaps		
OY	1889 aggaacatgtgacccagcagcatctctgtatctgttgcctgaagctacagaagaagcatttatt 1948		
DB	182 aggatatgtgtgactacagctctctacgcttcttggaagccacagaaaaagatttttt 241		
OY	1949 tcaaaaatgttgcatttttgatctctctgaacatggaagacaaagcctgactatgtgaagc 2008		
DB	242 tcaaaaatgtactatactaatctctggaattggaagaaaatctcctgaacaaaggc 301		
OY	2009 caaacctgagacgctacaacaaatgctgattgttctgttgcctgagctactactcctcaggta 2068		
DB	302 caaacatgtaaaacaacataaacatgtctgattgtatattgttgcaccacctcactccaggta 361		
OY	2069 atgaagaacccctacactgacgacgaatgaggccaactgtgagagagaagggtgaagaagatccacc 2128		

Db 362 gaatgaaacataacaaacagcttccagaaatgtagaagaagaggaataact 421
QY 2129 tcaactctgattcattgacggaanaaaagttaagctgaaatgagccaaagttag 2188
Db 422 tcaacctgacctctactctggaanaaaacaaatgaataatgagccaaagttag 481
QY 2189 ttgtcgaatgagtggtctacatcagatgtagagatattgaacgaataatgtaga 2248
Db 482 ttgtcgaatgagtggtctacatcagatgtagagatattgaacgaataatgtaga 541
QY 2249 aattctactatccaaatggaagaatac---aagcaatgaatgtagagagattactg 2305
Db 542 ctctctacgctgccaagtcacaaaacaaatcgaaacaaagtgtcccgagatctctg 601
QY 2306 gtaacaaatgtagtaaaagatgtagagagagagctgttccacaaagaatgtagatca 2365
Db 602 gtagaaatagagttataagtgtagaagagagagagctgttctagtagagcagtagaattg 661
QY 2366 ataaagttaacagagctctatgaaagaagatgtagttgttctccaaatcccgagagag 2425
Db 662 attctacaaacaaactgtatgaaagaatgtagaatacttctctgataaagtaacaaag 721
QY 2426 agaaagcttataaagttagtcacaaatgtagatctatagttgtaattctgtacagaa 2485
Db 722 aaaaagatcccaatagtttataatgcaaaagtattgtatctgtgttgaattgttaacgaa 781
QY 2486 aaaaacaaacaaagaagctcccaacaaagaacaaataaataatcctcgaagacat 2545
Db 782 aaaccttaatacagaagctcccaagcctacaaacaaataaagttagaatttagaagtaac 841
QY 2546 gtagaagttagctgtatctctgagagactttaagaacaaacctctatgacaaacagac 2605
Db 842 gtagaagttagctgtatctctgagagactttaagaacaaacctctatgacaaacagac 901
QY 2606 caatccacactctcaatgtctgacagattggaagaagaattgtgtttagtcttgaca 2665
Db 902 ctccacactgtctctcattgtctgaaagatcagtcacaaagaattgtgtttagtcttgaca 961
QY 2666 aatctggaagcagctgagactgtgaacgcctcaatcgaactgaaatcagagccagact 2725
Db 962 agctctggaagcagctgagactgtgaacgcctcaatcgaactgaaatcagagccagact 1021
QY 2726 tctgtctgacagagctgtgagctgtgagctgtgagctgtgagctgtgagctgtgag 2785
Db 1022 tctgtctgacagagctgtgagctgtgagctgtgagctgtgagctgtgagctgtgag 1081
QY 2786 ccacatgataacaaatgacatatacagataaacagtgtagtgaacagagacacactcgca 2845
Db 1082 ccacatgataacaaatgacatatacagataaacagtgtagtgaacagagacacactcgca 1141
QY 2846 aaagattacctgacagagcttccagagagagagctcagctgacagagcttccagctgac 2905
Db 1142 caagattacatcatctctctgtaggaagacttccatctgctgtaataatgacat 1201
QY 2906 ttactgtatgtgacacattgtcctgttccatgacaaacagagttatgtagagagctg 2965
Db 1202 ttcaagtgtagtagagagc----- 1219
QY 2966 gacaagaataatccaattgagcctctctgagcctgagcttagtagatagaagaataatc 3025
Db 1220 -----tacattccc 1228
QY 3026 caactgtagatctgaatgtgtctgtctgaagagatgtaggaagaacacatataatgagct 3085
Db 1229 aactcgatgtagctcgagagtagctgtctgtactgtgtaggagtagaacacagagcttct 1288
QY 3086 gctttaaagagtgcaacaaagtgtgtgcataatccacacagctgctgtgtaggagcctctg 3145
Db 1289 gtagttagtagagtgaaacaaagtgtgagcattgtcatattgttctgtgtagagagctg 1348
QY 3146 cagcttaagaacatcagaagagctgtccaaaatgacagagagtttaacagacatagcttcag 3205

Db 1349 ctgataagcagtaataatagatagtagaagataacagaggaagtagtattttagtctcag 1408
QY 3206 atcaagttcagaacaaatgagctcattgagcttctgtgagcccttcacatagaagaatgag 3265
Db 1409 atgaagctcagaacaaatgagctcattgagcttctgtgagcccttcacatagaagaatgag 1468
QY 3266 ctgtctctcagctccatccacatcagctgtgagagtaaggatlaaacctccagaacagcag 3325
Db 1469 atctctccagaagagctccatccacatcagctgtgagagtaaggatlaaacatgtagtagcct 1528
QY 3326 gtagttagtagcagaatgagctgtgagagacacagtggaagaagagacattgttcttata 3385
Db 1529 gtagttagtagcagaatgagctgtgagagacacagtggaagaagagacattgttcttata 1588
QY 3386 cctggaacatgtagcagctcccaacatcctctctgagagatccagtgtagcagaagcagctg 3445
Db 1589 catggaacagctgctcctccagatattctctctgagagatccagtgtagcagaacatgtaga 1648
QY 3446 gctttagtagtagcagaacaaatgtagcagcttccatccacatccagcagctgtcagag 3505
Db 1649 attctcagtagtagtagcagacttccaaatgagcctatctcagtagtccagagagctcagag 1708
QY 3506 ttgagcactggaataatcagctgtgc-----aagcaagctcacaacacttgagcctgagc 3559
Db 1709 ttgagcactggaataatcagctgtgc-----aagcaagctcacaacacttgagcctgagc 1768
QY 3560 tcaagctcagctgtgtagcagcttccatccatccatccatccatccatccatccatccatcc 3619
Db 1769 taactctctgagagagaaatctctctgtctcctcacaacagtagtagtgaataatgtaga 1828
QY 3620 aggaacacagcaaatccccaagcctctgtgtagttagttagttagttagttagttagttag 3679
Db 1829 aggaacacagcaaatccccaagcctctgtgtagttagttagttagttagttagttagttag 1888
QY 3680 cccaactctcagagagcagctgtcagagcctgtgtagttagttagttagttagttagttag 3739
Db 1889 taactctctgagagagaaatgtgagccttcaatgtagcagagagtagcagagagtagcag 1948
QY 3740 ccttgaactactgtagtaatgtagcagagctgtgtagttagttagttagttagttagttag 3799
Db 1949 ttgtgagactttgtagtaatgtagcagagagtagcagagagtagcagagagtagcagagag 2008
QY 3800 caagtagtattcaactatgacagaaatgtagcagagagtagcagagagtagcagagagtag 3859
Db 2009 ccaagtagtattcaactatgacagaaatgtagcagagagtagcagagagtagcagagagtag 2068
QY 3860 gaagtagttagcagagagtagcagagagtagcagagagtagcagagagtagcagagagtagc 3919
Db 2069 gaagtagttagcagagagtagcagagagtagcagagagtagcagagagtagcagagagtagc 2128
QY 3920 ctgagctgagtagagagtagaataacatgtagatcccaagagcctgaaatgaataag 3979
Db 2129 cagagctgagtagagagtagaataacatgtagatcccaagagcctgaaatgaataag 2185
QY 3980 atgagtagttagaacaacagcagagtagttagcagagagtagcagagagtagcagagagtag 4039
Db 2186 agtagtagtagcagagcagctgtgagagtagttagcagagagtagcagagagtagcagagagtag 2245
QY 4040 ctctgagtagtagcagagcagctgtgagagtagttagcagagagtagcagagagtagcagagagtag 4099
Db 2246 tatcaagaatgtagcagagccttccctgtgagcagacatacccaagtagcagagagtagcag 2305
QY 4100 tgaagcaggaataatcagagagtagcagagtagcagagtagcagagtagcagagtagcagagtag 4159
Db 2306 ttgagtagcagagtagcagagtagcagagtagcagagtagcagagtagcagagtagcagagtag 2362
QY 4160 attagtagcagagtagcagagtagcagagtagcagagtagcagagtagcagagtagcagagtag 4219
Db 2363 atttagtagtagcagagtagcagagtagcagagtagcagagtagcagagtagcagagtagcagagtag 2422
QY 4220 tgaagtagcagagtagcagagtagcagagtagcagagtagcagagtagcagagtagcagagtag 4279
Db 2423 taagtagcagagtagcagagtagcagagtagcagagtagcagagtagcagagtagcagagtagcag 2482

OY	4280	ccaactctggggaagtccttttggtttaaccocgaanaacattcctttggaatgtgcag	43399
Db	2483	ccaaactcaaggagaagtcttgtagcatltaaacaccagaataatctcagaagaataatgcaccc	25472
OY	4340	atcttccatcgacatccagcgtctgtgataaggtcgatctggaatcgaatatccaca	43399
Db	2543	acataattatttgcatttaaaaatgatagataaagcaatttgcattcaataaagtatccaca	26072
OY	4400	ttagcacgagatctcttggtttatctctccacagaactccgccagaga--cacctaagtctgy	44562
Db	2603	ttagcacagaactcttggtttatccctcacaagcaaatcctgatgacattgaccttacacca	26652
OY	4457	atgaaacgtctgctcctctgctccacaata	4483
Db	2663	ctctactcctactcctcctactccctcgata	2689
<hr/>			
RESULT_15			
ID	AAAF44241	standard; cDNA; 3265 BP.	
XX	AAF44241;		
DT	02-APR-2001	(first entry)	
DE	Human PR01124 (UNO562) nucleotide sequence SEQ_ID NO:378.		
XX	Human:	secreted and transmembrane protein; PRO; cytosolic;	
KW	cell death; cancer; chromosomal mapping; gene mapping; tissue typing;		
XX	diagnostic assay; ss.		
OS	Homo sapiens.		
XX	WO200073454-A1.		
PN	07-DEC-2000.		
PD	30-MAR-2000; 2000MO-US08439.		
XX			
PF	02-JUN-1999; 99MO-US12252.		
PR	23-JUN-1999; 99US-O141037.		
PR	07-JUL-1999; 99US-O143048.		
PR	20-JUL-1999; 99US-O144758.		
PR	26-JUL-1999; 99US-O145698.		
PR	28-JUL-1999; 99US-O146222.		
PR	17-AUG-1999; 99US-O149396.		
PR	15-SEP-1999; 99MO-US21090.		
PR	15-SEP-1999; 99MO-US21447.		
PR	08-OCT-1999; 99US-O158663.		
PR	30-NOV-1999; 99MO-US28313.		
PR	01-DEC-1999; 99MO-US28301.		
PR	16-DEC-1999; 99MO-US30095.		
PR	20-DEC-1999; 99MO-US30911.		
PR	05-JAN-2000; 2000MO-US00219.		
PR	06-JAN-2000; 2000MO-US00376.		
PR	11-FEB-2000; 2000MO-US03655.		
PR	18-FEB-2000; 2000MO-US04341.		
PR	22-FEB-2000; 2000MO-US04414.		
PR	24-FEB-2000; 2000MO-US04914.		
PR	24-FEB-2000; 2000MO-US05004.		
PR	02-MAR-2000; 2000MO-US05841.		
PR	15-MAR-2000; 2000MO-US06684.		
PR	20-MAR-2000; 2000MO-US07377.		
XX			
PA	(GETH) GENENTECH INC.		
XX			
PI	Ashkenazi AJ, Baker KP, Bolstein D, Desnoyers L, Eaton DL,		
PI	Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski P,		
PI	Grimaldi CJ, Gurley AL, Kijavlin IJ, Napier MA, Pan J, Paoni NF,		
PI	Roy MY, Stewart TA, Tumas D, Watanabe CK, Williams JM, Wood WI,		
XX	Zhang Z;		

DR	WPI: 2001-032160/04.
DR	P-FSDB; AAB65272.
XX	
PT	PRO polynucleotides used to produce polypeptides used to target
PT	bioactive molecules such as toxins, radiolabels or antibodies, to
PT	specific cells, to cause targeted cell death -
XX	
PS	Claim 2; Fig 273; 935bp; English.
XX	
CC	The present invention describes human secreted and transmembrane PRO
CC	proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC	can be used for targeted delivery of bioactive molecules, such as
CC	toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC	sequences, and their fragments, can be used as hybridisation probes, in
CC	chromosomal and gene mapping, and in the generation of anti-sense RNA
CC	and DNA. They may also be used to produce transgenic animals which are
CC	used to develop and screen therapeutically useful reagents. The PRO
CC	nucleotide and protein sequence can be used for tissue typing and in
CC	treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC	AAPF4270 to AAPF4470 represent PCR primers and hybridisation probes used
CC	in the isolation of human PRO sequences. AAPF4087 to AAPF4265 and
CC	AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC	sequences given in the exemplification of the present invention.
SQ	
	Sequence 3265 BP; 1159 A; 596 C; 632 G; 878 T; 0 other;
Query Match	24.7%; Score 1129.8; DB 22; Length 3265;
Best Local Similarity	67.1%; Pred. No. 2e-300;
Matches 1749; Conservative	0; Mismatches 747; Indels 111; Gaps 6;
OY	1889 aggacatgttgcccccagcatctcgttacttcctttgaagctcacgaagaagcattttat 1948
Db	182 aggatcatgttgactaacagcttctaogtaaccttttgaagccacagaaaaaatltttt 241
OY	1949 tcaaaaatgtgcacatttgtattcctctgaacaatygaaagcaaaagtctactgtlyagac 2008
Db	242 tcaaaaatgtatcatataatttccttgagaatvtgaaaggaaaaatcctcgaccaaaagc 301
OY	2009 caaaccttgagaccctacaaaatgtcgtatgttctgtgttgcgtactactctccagga 2068
Db	302 caaaaatvgaaaacccaataacatgcgtatgtatagtttgacacacactaacctccagga 361
OY	2069 atgatatgccctcacacttgcagatvgycaactgttgaagaagaaggtgaaagatccacc 2128
Db	362 gggatgaaccatcacaccacaagcagttcacagaatbtgysaagaagcgatatactcact 421
OY	2129 tcaactcgtatttcattctcaggaaaaaagtlctgtuatatcatyagaccacaagtagggcat 2188
Db	422 tcaaccttacccttctactcttgaaaaaaaacaaattpaatatatygacacacagcaaatct 481
OY	2189 ttgtccatagttggyggtctacacgaatggggagatatttgacagttacaaataatgatgaga 2248
Db	482 ttgtccatagttggyggtctaccccggttggggagtttttgaatgagtcacatgaagatcagc 541
OY	2249 aattctacttatccaatyggaagaatac--aagcagttaagatyltcaagaggtatbatty 2305
Db	542 ctcttaccgltgtaagtcataaaaaaatcgaagcaacaagagtyttccgagytactctgy 601
OY	2306 gtacaaatgtagtaaagaagtgtlccaggsggcagctgttaacaccaaagaatgatcacttca 2365
Db	602 gttagaataagatlttaagtgttcaaggsgagcgttcttagttagatgcagatcgtg 661
OY	2366 ataagaatcacagagactctatgaaaaagagttgagtttcttccacatccgcagaagcy 2425
Db	662 attcttcaacaacaactcgtatggaaaaagatgttcaattcttcttcgtataagatracaaacy 721
OY	2426 agaaggtctctataatgtttgcacaaacatgttgatctctatagttgaattctytacagaac 2485
Db	722 aaaagaatcattcaatagtttatgcaaaagatttgatctctgttltgattttgaaagaaa 781
OY	2486 aaacacacacaaagaaggtccaaacaaagaataatcaaaaatgaaatcccgaaagcat 2545

QY	3620	aggagaccagcgaatctcccaagcccttcgtgtagttatgcaatatcttcgaagagcct	3679
Db	1829	aggagcctaaacagttcttcccaagcccaatgatactgttcaagcagaatctccaaagata	1888
QY	3680	cccaatctccagcgcagctgtccacagccctgtatgtaatcagtgatgtnaaacagta	3739
Db	1889	tactctgtcttgagcccaatgtgactgtcttcaattgaatcacagatgtgacatacaga	1948
QY	3740	cccttggaactactgataatgagacaggtgcgtgatactacataagatgacgtgtctac	3799
Db	1949	ctcttggaactcttgataatgagctgcagcgtctatcttccaaagaatgatagtgac	2008
QY	3800	caagatcttccaaactatgacaacagatgtgatacagctgttaaaagtgcggctctg	3859
Db	2009	ccagatctttaccagctataacagaaaatgycagataatgcttaaaagttcggctcag	2068
QY	3860	gagagatcaacgcacagccagacgagagtgatacccaagacagctgagacactgtacac	3919
Db	2069	gagagacaaacacgcagcgtctaataatgacgctccatctgataagacgcgtctacac	2128
QY	3920	ctgcgtgatctgagaaatgataatacaatggaatcccaagacgtgaataataag	3979
Db	2129	cagcgtcgtgtagtgaaacgagggaatctgaaacaaacccgcgaagacctgaaatgtc	2185
QY	3980	atgagtctcaacacaaacagctgtgttctcagcgaacatctccggagagctcatctg	4039
Db	2186	agagatctcagacacacttgagagatcttaagccgaacagcattcggaggttgatctg	2245
QY	4040	ctctctgatgtcccaaatgctcccaatcagatctctctccacccgcgaataacacgac	4099
Db	2246	tatcaacagtcctccaaagctctccctctgcgtgacaataacacgaagctcaatcacag	2305
QY	4100	tgaaagcggaaatcacagcgggagcagctccataatcttgactggaacagctccctgg	4159
Db	2306	ttgagtccacagctcatgagag---ataagtatcttcaatgtgacagacagagagta	2362
QY	4160	attatgacatggaacagctccacaagtatacttcgataaagtacaaagtatcttgatc	4219
Db	2363	attcttgatgtggaagaagcttcaacgtatatactcaagaataagctgcaagtatcttg	2422
QY	4220	tcgaagcgaagtccaatgaaatctctcaagtgatactactgctctcaatcccaagaag	4279
Db	2423	taagagacagtttgatgatagcctctccaagtaataactactgatactgtgcacccaag	2482
QY	4280	ccaacctcvggaagctctttgtttaaacacgaanaacttaactcttgaaatgycag	4339
Db	2483	ccaactccaagaaagcttgatcatcttaaacacgaanaatctccagaagaatgcaacc	2542
QY	4340	atcttctcatgtcatctcagcgtgtgataagttgatactgaaatccagaataaccaca	4399
Db	2543	acatatctatgtgcataaaagataagataaaagcaatcttgacatacaaaagtatacca	2602
QY	4400	ttgacagatatactctgttatctctccacagacatccgcgaagc---aacctagctg	4456
Db	2603	ttgcacaagaatcctgtttatccctccaaagcaatctcgtgacatgtacatccatacta	2662
QY	4457	atgaaacgtctgcctctcttccataa	4483
Db	2663	ctcctactccatctccactccactccggtata	2689

Search completed: March 30, 2002, 13:07:20
Job time: 21437 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 12:59:18 ; Search time 8456.85 seconds

(without alignments)
8912.966 Million cell updates/sec

Title: US-09-867-034-3

Perfect score: 4569
Sequence: 1 atcgataaattcctcaacac.....tcgagctgcaatagcctag 4569

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_ov:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rnd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rnd:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	2492.2	54.5	3311	6	AX193489	AX193489 Sequence
2	2492.2	54.5	3311	9	AF039400	AF039400 Homo sapi
3	2473.6	54.1	2826	9	AF127036	AF127036 Homo sapi
4	1847.2	40.4	2022	9	AK024970	AK024970 Homo sapi
5	1637.8	35.8	3079	4	AF095584	AF095584 Sus scrofa
6	1517.6	33.2	2937	10	AB017156	AB017156 Mus muscu
7	1136	24.9	3204	9	AF127035	AF127035 Homo sapi
8	1129.8	24.7	3221	9	AK000072	AK000072 Homo sapi
9	1058.2	23.2	3265	6	AX092338	AX092338 Sequence
10	1058.2	23.0	175591	2	AC068071	AC068071 Homo sapi
11	1051.8	23.0	140718	2	AS456270	AS456270 Homo sapi
12	1049.4	23.0	118426	2	HS431C21	HS431C21 Homo sapi
13	1033.8	22.6	196805	2	AC010275	AC010275 Homo sapi
14	1024.8	22.4	168185	2	AC079805	AC079805 Homo sapi
15	1005.4	22.0	149428	2	AC010264	AC010264 Homo sapi
16	1005.4	22.0	186882	9	AL136307	AL136307 Human DNA
17	1005.4	22.0	187045	2	AC073253	AC073253 Homo sapi
18	1004	22.0	151200	2	AC015668	AC015668 Homo sapi
19	1004	22.0	154712	2	AC010958	AC010958 Homo sapi
20	1003.8	22.0	99074	2	HS560B9	HS560B9 Human DNA s
21	1003.8	22.0	151551	2	AC021291	AC021291 Homo sapi
22	1003.8	22.0	152408	2	HS191J18	HS191J18 Human DNA
23	1003.8	22.0	157670	2	AC074331	AC074331 Homo sapi
24	1003.8	22.0	170425	2	AC024490	AC024490 Homo sapi
25	1003.8	22.0	176087	2	AC024625	AC024625 Homo sapi
26	1003.8	22.0	186780	2	AC005740	AC005740 Homo sapi
27	1003.8	22.0	187766	2	AC016042	AC016042 Homo sapi
28	1003.8	22.0	197927	2	CNS05TE0	AL158293 Human chr
29	1002.4	21.9	126039	9	AC012665	AC012665 Homo sapi
30	1002.4	21.9	168937	2	AC018747	AC018747 Homo sapi
31	1002.2	21.9	110000	2	AL360016_1	Continuation (2 of
32	1002.2	21.9	130244	9	AL590705	AL590705 Human DNA
33	1002.2	21.9	150011	2	AC068232	AC068232 Homo sapi
34	1002.2	21.9	157832	2	AC011653	AC011653 Homo sapi
35	1002.2	21.9	182938	2	AL512304	AL512304 Homo sapi
36	1002.2	21.9	196934	9	AL355530	AL355530 Human DNA
37	1002.2	21.9	197274	2	AL357373	AL357373 Homo sapi
38	1002.2	21.9	211544	9	AC025165	AC025165 Homo sapi
39	1000.6	21.9	70113	9	AC002478	AC002478 Human BAC
40	1000.6	21.9	122272	9	AL359845	AL359845 Human DNA
41	1000.6	21.9	139737	9	HS228H13	AL031985 Human DNA
42	1000.6	21.9	141320	2	AC074322	AC074322 Homo sapi
43	1000.6	21.9	148098	2	AL359315	AL359315 Homo sapi
44	1000.6	21.9	150983	2	AL138802	AL138802 Homo sapi
45	1000.6	21.9	159160	2	AC018624	AC018624 Homo sapi

ALIGNMENTS

RESULT	1					
AX193489	AX193489	3311 bp	DNA			
LOCUS	Sequence 1056 from Patent WO0149716.			PAT	15-AUG-2001	
ACCESSION	AX193489					
VERSION	AX193489.1	GI:15211440				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
AUTHORS	Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J., Stoilk,J.A., King,G.E., Wang,T., and Jiang,Y.					
TITLE	Compounds for immunotherapy and diagnosis of colon cancer and methods for their use					
JOURNAL	Patent: WO 0149716-A 1056 12-JUL-2001;					
FEATURES	CORIXA CORPORATION (US)					
source	1. 3311 Location/Qualifiers					


```

Oy 3575 ccaatgctacccctgcctccaattacagtgaactcccaaaacgaagaacacgaacgaat 3634
    |||||||
Db 2102 ccaatgctacccctgcctccaattacagtgaactcccaaaacgaagaacacgaacgaat 2161
Oy 3635 tcccaagccctcgtgtaattatgtaaatatcgccaaggagcccccgaattccagg 3694
    |||||||
Db 2162 tcccaagccctcgtgtaattatgtaaatatcgccaaggagcccccgaattccagg 2221
Oy 3695 ccaagtctacagccctgtaattgaatcagtgaaataacagttacacttggaactctg 3754
    |||||||
Db 2222 ccaagtctacagccctgtaattgaatcagtgaaataacagttacacttggaactctg 2281
Oy 3755 ataattgagcaggtgctgtaactgaagatgacggtgctactcaagtaattcaca 3814
    |||||||
Db 2282 afaattgagcaggtgctgtaactgaagatgacggtgctactcaagtaattcaca 2341
Oy 3815 ctcttgacagaatggtgataacagtgtaaaagtgcggctcttggaaggatgaacg 3874
    |||||||
Db 2342 ctcttgacagaatggtgataacagtgtaaaagtgcggctcttggaaggatgaacg 2401
Oy 3875 ccaagcaggaagatgatacccaagcagatggaacactgatactgctggaattgaga 3934
    |||||||
Db 2402 ccaagcaggaagatgatacccaagcagatggaacactgatactgctggaattgaga 2461
Oy 3935 atgatgaaatacaatggaatcccaacgaacctgaataataaggaatgatttcacaca 3994
    |||||||
Db 2462 atgatgaaatacaatggaatcccaacgaacctgaataataaggaatgatttcacaca 2521
Oy 3995 agcaagtgcttgaagcagaacatcctcggaagctcatttggtcttctgataccaa 4054
    |||||||
Db 2522 agcaagtgcttgaagcagaacatcctcggaagctcatttggtcttctgataccaa 2581
Oy 4055 atgtcccaatccctgatactcttcccaactggaacacacgaactggaagcggaattc 4114
    |||||||
Db 2582 atgtcccaatccctgatactcttcccaactggaacacacgaactggaagcggaattc 2641
Oy 4115 acgggggagcagtcataactcagtcagtcagtcctcggaagatgatatgacatgga 4174
    |||||||
Db 2642 acgggggagcagtcataactcagtcagtcagtcctcggaagatgatatgacatgga 2701
Oy 4175 cagtcacaatataatacttgaaataagtaaatatcttctgatactcagaagaagtcca 4234
    |||||||
Db 2702 cagtcacaatataatacttgaaataagtaaatatcttctgatactcagaagaagtcca 2761
Oy 4235 atgaatcctcagaatataactgctcctcatcccaagaagaacgaactctgagaag 4294
    |||||||
Db 2762 atgaatcctcagaatataactgctcctcatcccaagaagaacgaactctgagaag 2821
Oy 4295 tcttttcttaaacacgaataacttacttgaataatgagacagatcttctcatgcta 4354
    |||||||
Db 2822 tcttttctttaaaccacgaataacttacttgaataatgagacagatcttctcatgcta 2881
Oy 4355 ttcaaggcgttgataagtgatcgaatcgaataatgaataatccaaactctgacgaagtatct 4414
    |||||||
Db 2882 ttcaaggcgttgataagtgatcgaatcgaataatgaataatccaaactctgacgaagtatct 2941
Oy 4415 ttttatatccctcagaactccgccaagagacactcagatcagtaataaagtcctgctct 4474
    |||||||
Db 2942 ttttatatccctcagaactccgccaagagacactcagatcagtaataaagtcctgctct 3001
Oy 4475 gtctcaatatataatacagaacacactcctctgcatcacaatttaaaatlatgta 4534
    |||||||
Db 3002 gtctcaatatataatacagaacacactcctctgcatcacaatttaaaatlatgta 3061
Oy 4535 agtggataggaagcagtcagtcgttcaataagcctag 4569
    |||||||
Db 3062 agtggataggaagcagtcagtcgttcaataagcctag 3096

```

```

RESULT 2
LOCUS AF039400 3311 bp mRNA PRI 14-DEC-1998
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA,

```

```

ACCESSION complete cds.
AF039400
VERSION AF039400.1 GI:4009457
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 311)
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreier,K.D., Fuller,C.M. and
Pauli,B.U.
TITLE Genomic cloning, molecular characterization, and functional
analysis of human CLCA1, the first human member of the family of
Ca2+-activated Cl- channel proteins
JOURNAL Genomics 54 (2), 200-214 (1998)
MEDLINE 99047526
REFERENCE 2 (bases 1 to 311)
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,
Cornell University, Ithaca, NY 14853, USA
FEATURES
Source
1..3311
Location/Qualifiers
1..3311
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p22-p31"
/tissue_type="small intestine"
1..3311
/gene="hCLCA1"
/note="transmembrane glycoprotein"
/codon_start=1
/product="calcium-dependent chloride channel-1"
/protein_id="AAC95428.1"
/db_xref="gi:4009458"
/translacion="MGPERSVFLILHLLEGALSNSLIOLNNNGYREGIVAYDNPV
EDFLIOIKDMVMOASLYLEFAGKRRYEFNVAILPEPMKTRADYVRPLERYKKA
DVLVAESTRPNDPPTROMKNGCEBERHILTPDFLAGKLAEYPOGKAFVEMNH
LRMGVFDYNDKERFVLSNGRIQAVRCSAGITGVNVKCGGSGCYTRCFNVTGL
YKGECEFLVNSRQTEKASIMPAQVDSIVERCTQDNHKKKPNQONOKNLRSTVEYI
RDSDFPKTTPMTQPPNPFPFLSLQIGRIYCLVLDKSGSATNRLRLNQAQLEL
LOVELSGWMTFEDSAAHVQSELIQINSQSDRDTLAKRLPAASGGSTISGSLRS
FTVIRKKYPTNGSEIVLITDGEEDITSGCFEWEKOSCALIHTVLAQPAQDELISK
MTGSLQRTYASQVONNGLIDFGLSSNGAVSQRSIOLESKITLQNSQNMETIVY
DSIVGKDTFLITWTTPQILLMDPSQKQGFVVDKNTMAVLIQIPGAKVETWYX
SLQASQTLTLTVTSRASNAFLPIITVYSKINKDTSKEPPLVYVYANRQOASPLRA
SVTALIESVNGKVTYLELDNGAGADATKDDGVSRFTYDNGRYSVKRALGVN
AARRVAPIQOOGALYIPGMILENDEIOMNPPEPEINKDOVHKQVCFSTSSGSEVAS
DVNPAPIPDLEPPQIDIDKAEIHGSLINTWTAQPDGDVGHAAHYIIRISYIID
LRQFNSGLQVNTTALIPKRAKSEFVFLFKENTPTFRNGDTLFAIQAVDVKSEI
SNIRVSLFIPQTPPETPSPDETSAQPCNPHINHSITPIGHILKIMMKWIGELQSLIA
BASE COUNT 1028 a 692 c 742 g 849 t
ORIGIN
Query Match 54.5%, Score 2492.2; DB 9; Length 3311;
Best Local Similarity 91.3%; Pred. No. 0;
Matches 2772; Conservative 0; Mismatches 8; Indels 255; Gaps 2;
Oy 1535 tggccaagaagaataacacaggaagatgatacagaagaatgggcacatttaagatctgtgt 1594
    |||||||
Db 317 TTCAAAAGAGAGAAATCACAGGAGAGATGTACAGCAATGGGCCATTAAAGATCTGTGT 376
Oy 1595 tcaatctatcttcaactctagaaggagccctgagtaattcaactcaatcaagtcagaca 1654
    |||||||
Db 377 TCATCTTATTTCTTACCTTTAGAGAAGGCCCTGAGTAATTCACACTATTACGCTGAACCA 436
Oy 1655 acaatgcatgaaggaatgctgcttgcaatcgaccccaatgctgcagaagaatgaacac 1714
    |||||||
Db 437 ACAATGCGTATGAAGGCAATGTGCTGCAATCGAACCCCAATGTGTGCAAGATGAACAC 496

```

QY 1715 tcaatcaacaataaaggaggagtagacacgctcaacaaga tgaaggaggagagtcagagaga 1774
|||||
Db 497 TCATTCAACAATA----- 510
QY 1775 aactctctctcccgctcaataatacatacaacacacacgacgacaagctcgtgtgc 1834
511 ----- 510
QY 1835 acaacaacagcccatgacacacagcagacatatacgcgacacacacgacgctcagagaga 1894
511 -----AAGACA 517
QY 1895 tggtagaccaggaactcctgtlatctgtctgaagctacaggaagcgaatttattcca 1954
|||||
Db 518 TGGGAGCACCAGGCACTCTGTATCTGTTTGAAGCTACAGAAACGATTTATTTCAAA 577
QY 1955 atgttgcatttgatctccgtaaacatgaaagacaaagcgtgactatgtgagaccaaac 2014
578 ATGTTGCCATTGTGATTCGTAACATGGAAGACAAAGCGCTGACTATGTGACCAAAAC 637
QY 2015 ttgagacctcaaaaatgctgagtgtctgtgtgtgactactccctccaggtatgtg 2074
|||||
Db 638 TTGAGACCTACAAAATGCTGATGTCTGTGCTGAGTCTACTCTCCAGTATATG 697
QY 2075 aacctacactgagcagatggcgaactgtgagagaaaggtgaaagga tccacctc 2134
698 AACCTACACTGACGACACATGGGCAACTGTGAGAGAAAGGTGAAGGATCCACCTCACTC 757
QY 2135 ctgattcattctgcagaaagaaagttagctgaatatgacccaaggaagtcaggtcctc 2194
758 CTGATTCATTGTCAGGAAAAAGTTACTGAAATGTGACCCCAAGTAAGGCAATTTGTC 817
QY 2195 atgagtgagcctcactcctcgatggaggagatttgaacggtacaataatgataagaattct 2254
818 ATGAGTGGCTCATCTAGATGGGAGTATTGTGACGATCAATAATGATGAGAAATCT 877
QY 2255 acctatccaatgagaagataccaagcaglaagatgtctcagcaggtatctatgtlaacaatg 2314
878 ACTATCAATGGAAGATACAAACAGTAAGATGTTCAGCAGGTATTACTGTCAAAATG 937
QY 2315 taagtaagaagtgctcagggagagcagctgttaccacaaagatgcacattcaataagtaa 2374
938 TAGTAAGAAAGTGTTCAGGAGGAGCTGTTCACCAAAAGTGCACATTCATTAAGTTA 997
QY 2375 caggacactatgaaagagatgtgagtgtgtctccaatcccgccagcagcagagaagctt 2434
998 CAGACCTCTATGAAAAAGATGTGAGTTGTCTCCAAATCCCGCAGAGGAGAAAGCTT 1057
QY 2435 ctataatgtttgcaacaacatgttgaattctatagttgaattctgtacagaacaacaca 2494
1058 CTATAAATGTTTGCACAAACATGTTGATTTGTTGTAATTCGTGACAGAAACCAACACA 1117
QY 2495 acaagaagaagctccaacaagaacaatacaaaaatgcgaatctccgaagcagatggaaatgta 2554
1118 ACAAGAAGCTCCAAACAAAGCAAAATCAAAATCAATCTCCGAGAGCAACAGGGAAGTGA 1177
QY 2555 tccgtgattctgaagacctttaagaaacacactcctatgacaacacagccacacaaatcca 2614
1178 TCCGTGATTTGAGGACTTTTAAGAAAACACTCCTATGACAAACAGCCACCAAAATCCCA 1237
QY 2615 ccttcattgtctgacagatgagacaagaatgtgtgtttagtccctgacaacatctgaa 2674
1238 CCTTCTCATTTGCTGCAGATTGACAAAGAAATGTGTGTTTGTCTTGACAAATCTGGA 1297
QY 2675 gcatggagcagtgtaaccgctcaatcgactgaaatcaagcagccagcttctcgtctgc 2734
1298 GCAATGGAGCAGGTAAACCGCTCAATGCAATGAATCAAGCAGGCGACCTTTTCTGTGTC 1357
QY 2735 agacagctgagcctgaggctccctgggtctggagatgtgacattggaacgtgcctgcatlac 2794
1358 AGACAGTTGAGCTGGGCTCTGGGTTGGGATGTGACATTTTGACAGTGTGCTGCCATGTAC 1417

QY 2795 aaaaatgactcatatacagataaacaagtgycagtgacagggacacacgctcccaaaagatlac 2854
1418 AAAAGTAACTCATATACAGATTAACAGTGCAGTGACAGGACACACTTCGCAAAAGATTAC 1477
QY 2855 ctgcagcagcttcaaggaggagcgtccatctgcagcggggtctgatatgcatlaactgata 2914
1478 CTCGAGCAGCTTCAAGGAGGAGCGCTCATCTGACGCGGCTTCATCGGCATTTTACT--- 1533
QY 2915 tglgacacattgucgtgttccatgacacacagcagttatggggagtgcgacaagaa 2974
1534 ----- 1533
QY 2975 atccaatlggcccctcgtgcctgcagcttagtgatgaagaanaatatccaactgag 3034
1534 -----GTGATTAGGAAGAAATATCCAACTGATG 1561
QY 3035 gatctgaattgtgctgtgctgacgagtaggggaagacaaactataagtgggtgtttaaag 3094
1562 GATCTGAATTTGTCTGCTGACGAGATGGGAGACAACTATATAGTGGGTCTTTAACG 1621
QY 3095 aggtlcaacaagaatggtgctcatatccacacagtcgcttggggccctcgcagctcaag 3154
1622 AGGTCAACAAGTGGTGCATCATCCACACAGTCCGTTTGGGGCCTCTGCAGCTCAAG 1681
QY 3155 aactagaagagctggtcccaaaatgacaggaaggtttacagacatatgcttcagatcaagtlc 3214
1682 AACTAGAGAGCTGTCCAAATATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTTC 1741
QY 3215 agaaacaatggccctcatatgagcttggggcccttcaatgaagaatggagcgtctctc 3274
1742 AGAACAAATGGCTTCATGATGATGCTTTTGGGCTTCATATGAGAAATGAGCTGTCTCTC 1801
QY 3275 agcgtcctatccagctgtgagatgaagatlaaacccctccagaacagcgaatgagatgag 3334
1802 AGCCCTCATCTCAGCTTGAAGATGAAGGATTAACCCCTCCAGAACAGCATGATGATG 1861
QY 3335 gcaacagtacgtgagacagacagcgtgggaaggaacactgttcttcatlacactgagcaa 3394
1862 GCACAGTATGATGTGACAGCAGCAGTGGGAAAGACACTTGTTCATACCTGTGCAAA 1921
QY 3395 tgaagcctcccaaatctctcgtggga tccagtggaagaagcagaagtggtgtttag 3454
1922 CGCAGCTCCCAAAATCTTCTGTGGGATCCAGTGGACAGAAAGGAGGCTGTGTAG 1981
QY 3455 tggacaaaacacccaatgagccttaccctcaaatcccaagcagatgtgaaagttgagcact 3514
1982 TGGACAAAAACCAAAATGAGCCCTACTCCAAATCCAGGCAATGCTAAGGTTGGCACTT 2041
QY 3515 ggaataacagtcgcaagcagactacaaacttgacctgtaactgcaagtcctccgtgct 3574
2042 GGAATATACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGTACCTCCCGTCCGT 2101
QY 3575 ccaatgttaccctgacctccaattacagtgacttccaaaagaaacagacacagcaaat 3634
2102 CCAATGTACCTCTCCCAATTAACAGTACCTTCAAACGAACAGAACCAAGCAAAAT 2161
QY 3635 tccccaagccctggttagttatgacaata ttcgccaaggaagcctcccaatctcaggg 3694
2162 TCCCAAGCCCTGTGTGTTTATGCAAAATTTGSCCAAGAGAGCTCCCAATTCCTCAGGG 2221
QY 3695 ccaagtcacagccctgattgaatcagtgaa tggaaaacagtlactcttggaactactgg 3754
2222 CCAAGTTCACAGCCCTGATTAATCAATGATGAAAGAAACAGTTTACTTGGAACTACTGC 2281
QY 3755 ataatgagcaggtgtcgtgactactaagataaggtgtcactactaaggtatttcaaa 3814
2282 ATTAATGAGCAGGTGTGATGCTAATAGATGAGCGGTGCTACTCAAGGTATTTTCAAA 2341
QY 3815 ctataacagaaatggtagatacagtg laaaaatgcyggggtctcgggagaggtlaaacgag 3874
2342 CTATATGACAGATGTATGATATACAGTGTAAAGTGGCGGCTCTGGAGGAGATTAAAGCAG 2401
QY 3875 ccagaacggaagatgatacccccaagag tggagcaactgtatacatacctgctgattgaga 3934

```

Db 2402 CCAGACGGAGTGTATACCCAGACAGGTGAGCAGCTGTACATACCTGGCTGGATTGAGA 2461
Oy 3935 atgataaatacagaatggaatcccaagaagcgaataataagaatgagatggttcaacaca 3994
Db 2462 ATGATGAATATATAATGGATTCACAGACGTGAATATATAGATGATGATCAACACA 2521
Oy 3995 agcaagtgtgttcagcagaacacccctcggaggagctcatttctgagcttcgagtcacca 4054
Db 2522 ACCAAGTGTGTTTCAGACAACATCTCTGGAGGCTCATTGTGGCTTGTGATGTCACCA 2581
Oy 4055 atgtcccaatcctgacatctctcccaactcgtgccaatcccaacgaactgaaagcggaatcc 4114
Db 2582 ATGCTCCCATCTATCTCTCTCCACCTGCGCAATATCACCGACCTGAGAGCGGAATTC 2641
Oy 4115 aagggggagcagctcttaactcgaacttgagcagcctcttgaggatgataatgacaaatgaa 4174
Db 2642 AGGGGGGAGCTCTCATTAATCTGACTTGGAGAGCTCTGGGATGATTAATGACCATGAA 2701
Oy 4175 cagctacagaatatacattcgaataagfacagaatattcttgatctcagaagaatgca 4234
Db 2702 CAGCTCACAGATATATCATTCGAATAGTACAAGTATCTTGATCTCAGACACAACTTCA 2761
Oy 4235 atgaatctcttaagtgaaatcactgctcctcatcccaagaagcaactctgaggaag 4294
Db 2762 ATGAATCTCTTCAGATGATATCTAGCTCTCATCCCAAGAGCAACTCTGAGAGAG 2821
Oy 4295 tcttttcttaaacagagaacattactcttgaataatgacagatcttctcattgcta 4354
Db 2822 TCTTTTCTTTAAACAGAAACATTACTTTTGAATATGGACAGATCTTTTCATTGCTA 2881
Oy 4355 ttcaagctgttgataagtgatctgaaatcagaataatccaaatcagacagatctt 4414
Db 2882 TTCAGAGCTGTGATTAAGTGCATGCAATGCAAAATATCCAAATGACACGATCTT 2941
Oy 4415 tgttatctctcccaagactccgcaagagacactagctcgtatgaaacgtctgctcctt 4474
Db 2942 TGTATTATCTCTCCACAGACTCCGCCAGACACACTAGTCTGATGAAAGCTGTGCTCT 3001
Oy 4475 gtcccaatattcatatcaacagacacattcctgagcatctacattttaaataatagtga 4534
Db 3002 CTCCTAATATTAATATCAACAGACACCATCTGCGCATTTACATTTTAAATAATATGTGA 3061
Oy 4535 agtgaataggaagactgcagctgtcaatagcctag 4569
Db 3062 AGTGATAGGAGACTGCAGCTGTCAATAGCCTAG 3096

RESULT 3
AF127036 2826 bp mRNA PRI 10-Aug-1999
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCL)
DEFINITION mRNA, complete cds.
ACCESSION AF127036
VERSION AF127036.1 GI:4585468
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2826)
AUTHORS Agnel,M., Verma,T. and Culouscou,J.-M.
TITLE Identification of three novel members of the calcium-dependent
chloride channel (CaCC) family predominantly expressed in the
digestive tract and trachea
JOURNAL FEBS Lett. 455 (3), 295-301 (1999)
MEDLINE 99364503
PUBMED 10437792
REFERENCE 2 (bases 1 to 2826)
AUTHORS Agnel,M. and Culouscou,J.-M.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des
Carrieres, Rueil-Malmaison 92500, France

```

```

FEATURES
source
Location/Qualifiers
1..2826
/organism="Homo sapiens"
/db_xref="taxon:9606"
/lissue_type="small intestine; colon"
1..2826
/gene="CaCC1"
5..2749
/gene="CaCC1"
/note="bovine epithelial chloride channel homolog"
/codon_start=1
/product="calcium-activated chloride channel protein 1"
/protein_id="AAD25487.1"
/db_xref="GI:4585469"
/translation="MGPKSSVFLLIHLBELSLSNLIQLNNKYEGLVAIDNPV
EDELIIQIKDVAITDASLLEFATGRKRFKVAALILEPTWTKADYVPLKLETKA
DLVAESTPPEGDEPYTDOMGNGCKGKRIHLTPPIAGKRLAEYGPGRVHVMH
LRMGVEDEYNDEKFPYLSNGRIQAVRCAGSLITGVNVRKCGGSCYFRCTENKVTGL
YEKGEFVLQSRQTEKASIMFAOHVDSIYERCTEONHNKKAENKONKCNLRSTEWI
RDESEFKTTPPTTQPPNPTSLQIGRIYGLVDKSGSNATGRRLRLNAGOLF
LOTVRLGSWGVWPPDSAAHYOSEIQLNSGSDRPTLKRPLAASGCTSTCSGRSA
PTVIRKKTPTDGESEIVLITDGEDNTISGFNFVKOSGAIIRTVALLGPSADLEELSK
MTGGLOTAADQVONNGLIDAFGLSNGAVSORSIOLSKGLTLOMSOMNGTVIY
DSYKSDILFLITWTPPOILMDPSGOKGGEVVDKNTKMAVLOIPGIAVGTWKY
SLQASSOTLFLVTSRASNAITLPITVTSKTKTSKFPSPVAVANIRQSGPILRA
SVTALIESVNGKVTVELLDNGAGADATKDGVSRYRPTTYDNGRVSVKVRAIGLRA
AARRRVIOOGSALYIRPMIENDERQAMPREIKRDVQHKQVCFSTSGSGSVAS
DVNRPATIDPRPQITDLKAITHGSLINLTWIAAGDYDGHAKYIRISTSLD
LKDKNESLQVNTTALIRKANSESEVFLKPEPNTIFENGTDLEFLAIOAVDKVLSKI
SNIRAVSLFIPQRPETPSPDETSAFCPINHINSTIPGIIHKIMKIMKIGLOLSIA"
BASE COUNT 875 a 623 c 632 g 696 t
ORIGIN
Query Match 54.1%; Score 2473.6; DB 9; Length 2826;
Best Local Similarity 96.3%; Pred. No. 0;
Matches 2582; Conservative 0; Mismatches 4; Indels 96; Gaps 1;
Oy 1888 aagagacatggttaaccagagcatctctgtatctgcttgaaagctaacagaaagcatttat 1947
Db 164 AAGGACATGCTGACCCAGGACATCTGTATCTGTGAAGCTACACAGAAAGCGATTATAT 223
Oy 1948 ttcaaaaatggttcacatttgatctccctgaacatagaaagcaagcctacatagtga 2007
Db 224 TTCAAAATATGTTGCCATTTTGATCTCTGAAACATGGAACACAAAGGCTGACTATGTGACA 283
Oy 2008 ccaaaacttgagactacaaaatgctgattgtctgctgagctactactccaggt 2067
Db 284 CCAAAACTTGAGACCTACAAAATATGCTGATGTTGTTGCTGAGCTACTCTCCAGGT 343
Oy 2068 aatgatgaacctacacatgagatgagcaactgtgagaaaggtgaaagatcac 2127
Db 344 AATGATGAACCTCTACACTGAGCAGATGGGCAACTGTGGAGAGAGGATGAAGGATCCAC 403
Oy 2128 ctccactcctgattcatcttcaggaaaaagttagctgaaatagagaccacaagatagga 2187
Db 404 CTCACCTCTGATTTTCAATTCACGAAAAAGTTAGCTGATATGAGCACAAAGGTAGGCA 463
Oy 2188 ttgttccatgagatgagctatcatcagatlgaggagatatttgagagatlacataatgatgag 2247
Db 464 TTTGTGCATGAGTGGGCTCATCTACAGATGGGAGATTTTGACGAGATACATATGATGAG 523
Oy 2248 aaatctactatccaatggaagataacagcaagtaagatgttcagcaggtattactgct 2307
Db 524 AAATCTACTTATCCAAATGGAAGATATCAACGATGAAGATGTTCCAGCAGTATCTGCT 583
Oy 2308 acaaatgtatgaagaagatgctcagaggagcagcttlaacccaagaatagacattcaat 2367
Db 584 ACAAAATGATGTAAGAGTGTACAGGAGGACGCTGTTAACCAAAAGATGACATTCAT 643
Oy 2368 aaagtaacagagactatgaaaagatgtgattgttctccaaatcccgccagagag 2427

```


Dh 644 AAAGTACAGGACTCTATGAAAAAGATGTGATTTGTTCTCCAAATCCGCCACAGCAGAG 703
Oy 2428 aaggtctctaagtgttgcacaacatgttgattctatagttagattctgtacagaaca 2487
Dh 704 AAGGCTTCTATATGTTTGCACACATGTTGATTCTTAGTGAATTCGTACAGAAACAA 763
Oy 2488 aaccacacaagaagagcttccaaagaacaatlcanaaatgcaatctccgaagacata 2547
Dh 764 AACCAACAACAAGAGCTCCAAACAGCAAAATCAAAAATGCAATCTCCGAAAGACATGG 823
Oy 2548 gaagtgtccgtgtatcttgagactttaagaacaacatccatagacaacagcacca 2607
Dh 824 GAAGTGAATCCGTGATGTTGAGGACTTTAAAGAAAACCACTCCTATGACACACACCA 883
Oy 2608 aatccacacttcacatgtctgcagaatttgcagaagaatttggtttgaagccctgaca 2667
Dh 884 AATCCCACTTCTCATTTGCTGCGAGATTGGACAAAGAAATTTGTTGTTTAGTCTTGACAAA 943
Oy 2668 tctggaagcaltgagcagcttgaacgcctcaatcgactgaatcaagaagccagctttc 2727
Dh 944 TCTGGAAGCATGGGAGCTGGTAACCGCCTCAATGACTGAATCAAGCAGGCCAGCTTTTC 1003
Oy 2728 ctgctgagaagatgttgctgggttctgggttgggaatgtgacatttgaagtgtcgc 2787
Dh 1004 CTGCTGAGACAGTTGAGCTGGGGTCTGGGTGGGATGGGATTTGACAGTGTGCC 1063
Oy 2788 catgtacaatgaactacacataaacaagtgcagtgacaggaacacacatcgccaaa 2847
Dh 1064 CATGTACAAAGTGAACCTATACAGATTAACAGTGGCAGTGCACAGGACACACTTGGCCAAA 1123
Oy 2848 agattacctgcagcagcttcaagaaggagcttcaatctgcagcgggtcttgcagcatt 2907
Dh 1124 AGATTACCTGCGAGCAGCTCAGAGGAGCGTCATGACCGGGCTTCGATCGGCATTT 1183
Oy 2908 actgatattgaggacaacttggctgtttccatgacacacagcagtcattggggagtgca 2967
Dh 1184 ACT----- 1186
Oy 2968 caagaanaatccaatltggcctctctgctgacagcttagtgaatgaagaatatcca 3027
Dh 1187 -----GTGATTAGGAAGAAATATCCA 1207
Oy 3028 actgattgatttgaattgtgtgtcgtgaacgaatggggaagaacaacataaagtgggtgc 3087
Dh 1208 ACTGATGATCTGAATTGTGCTCTGACGGATGGGGAACACACATTAAGTGGGTGC 1267
Oy 3088 tttaacagagtcacaacaagtgtgtccatcatccacaagtcgcttggggcctctgca 3147
Dh 1268 TTTTAACGAGGTCAAAACAAAGTGTGTCATCATCCACAGTGGCTTTGGGGCCTCTGCA 1327
Oy 3148 gctcaagaactagagagctgttccaaaatgacagaggtttacagacatatgtctcagat 3207
Dh 1328 GCTTAACAACTAGAGAGCTGTCCAAATGACAGAGGTTTACAGCATATATGCTTCAGAT 1387
Oy 3208 caagttcagaacaatggcctcaatgtatgtcttggggccttccatcagaagaatlgagct 3267
Dh 1388 CAAGTTCAGAAACATGGCCTCATTTGATGTGTTGGGGCCTTTCATGAGAAATGAGACT 1447
Oy 3268 gtctctaagcgtccatccagcgttgagaagtaaggatlaaacctccagaacagcagatgg 3327
Dh 1448 GTCTCTAGCGCTCATCCAGCTTGAAGTAAGGATTAAACCTTCAGAAACAGCAGTGG 1507
Oy 3328 atgaatgacagatgtatcgttgcagcagcacgttggaaagacacttgttctctatcac 3387
Dh 1508 ATGAATGGCAGTGTGATCGTGACAGCAGCGTGGAAAAGGACACTTGTTCCTTATACCC 1567
Oy 3388 tgaacaatgacgcttccccaatctctctbgygaatcccaatgagacagaagcaagtggc 3447
Dh 1568 TGGACACAGCAGCTCCCAAAATCCTCTCTGGAGATCCAGATGAGACAGAAAGCAAGTGGC 1627
Oy 3448 ttgtgtagtgaacaaaacacaaaatgtgcttactctccaatccaggcatgtcctaagttc 3507
Dh 1628 TTTGTAGTGAACAAAACACCAAAATGCGCTTACCTCCAAATCCAGGCAATTTGCTAAGGTT 1687

Oy 3508 ggcacttgaataatacaagtcctgcaagcaagctcaaaaacttgaacctgactgtcagctcc 3567
Dh 1688 GGCACCTTGGAAATACATCTGCAAGAGCTCACAAACCTTGACCTGACACTGTCACCTCC 1747
Oy 3568 cgtgcgtccaatgtctaccctgtcccaattacagtgacttccaaaacgaagaagaccc 3627
Dh 1748 CGTGCCTCAATGCTTACCTGCTCCCAATTAAGTGACTTCCAAAAGAAAGAGACACC 1807
Oy 3628 agaaaatccccagccctcgtgtagttatgcaaatatctgcagaagagcctcccaatt 3687
Dh 1808 AGCAAAATCCCAAGCTCTGTGATGTTATGCAAAATTTGCCAAGGAGCCTCCCAATT 1867
Oy 3688 ctcaaggccagtgctcacagccctgattgaaatcagtgatgtgaaacaaagttacttggaa 3747
Dh 1868 CTCAGGSCCACTGTACAGCCCTGATTGATTCAGTGAATGGAATAACAGTACTCTTGAA 1927
Oy 3748 ctactgataatgagcaggtgtcgtatgtactaaggaatgacgggtgtactccaaggtat 3807
Dh 1928 CTACTGATTAATGAGCAGGTGCTGATGCTAATGAGATGACGGTGTCTCAAGGTAT 1987
Oy 3808 ttcaacaattatgacagcaatgtgtatagatgttaaaatgctggggtctgggagagtt 3867
Dh 1988 TTCAACAATTATGACACGAATGGTAGTACAGTGTAAAGTGCAGGCTGTGGAGAGTT 2047
Oy 3868 aacgcagccagagagatgataccccagcagatlggaagcactgtacatactgtgtcg 3927
Dh 2048 AAGCCACCCAGAGCAGAGATGATACCCAGAGATGAGACAGTACATACCTGGCTGG 2107
Oy 3928 attgagaatgtgaataacaaatgaatccaaagacctgaataatgaagtatgtc 3987
Dh 2108 ATTGAGATATGAAATACATGAAATCCAAAGACCTTAAATTAAGATGATGATTT 2167
Oy 3988 caacacaagaagaatgtgttcaagacagacatctccgggaagctaatgttggctctgat 4047
Dh 2168 CAACACAAGAGTGTGTTTTCAGAGAACATCTCTGGGAGGCTCATTTGTGGCTTGAT 2227
Oy 4048 gtcccaaatgtcccatactgtacttctccaaacttggccaatlcacagacactgaagcg 4107
Dh 2228 GTCCCAATGTCTCCATACCTGATCTCTCCACCTGGCCAAATCACGACCTGAAGGCG 2287
Oy 4108 gaaatcacggggagcttccatlaatctgaacttgaagcagctctgggagatgattatgac 4167
Dh 2288 GAAATTCACGGGGCAGCTCATTAATCTGACTTGGACACTCTCTGGGATGATTATGAC 2347
Oy 4168 catggaacagctcacaagtalalcatcogaataagtaagaatttcttgcacagaagac 4227
Dh 2348 CATGGAACAGCTCACAGTATATTCGANTATGATCAAGTATTTCTTGATCTCAGAGAC 2407
Oy 4228 aagttcaatgaatcttctcaagtlgaatlaactgtcttccatcccaagaagccaactct 4287
Dh 2408 AAGTTCAATGAATCTCTCAAGTAATACCTCTCATCCCAAGGAAGCCAACTCT 2467
Oy 4288 gaggaaagcttltgtttaaaccagaacaacttaactttaaataatggcagaatcttctc 4347
Dh 2468 GAGGAAGTCTTTTGTGTTAAACCAAGAAACATTACTTTGAAAATGGCACAGACTTTTC 2527
Oy 4348 attgacttcaagcgtgtgttaagtgatcgtgaatcagaataatcaaatcatatgcaga 4407
Dh 2528 ATTGCTATTCAAGCTGTGTGATTAAGTGTGATCTGAATTCAGAAATATCCAAATTTGCACGA 2587
Oy 4408 gtaacttgttlatcttccacagactcgcagagacactagttctgtatgaaagctct 4467
Dh 2588 GTATCTTTGTTTATCTCCACAGACTCCGCGAGAGACACTAGTCTGATGAAGACGTCT 2647
Oy 4468 gctcctgtcttaatatlatcatatcacagacacaatctctggcatlcaatlttaaaatc 4527
Dh 2648 GCTCTTGTCTTATATTCATATCAACAGACCACTTCTGCAATTCACATTTAAATAAT 2707
Oy 4528 atgtggaagtgttaggagaagcagctgtcaatagccag 4569
Dh 2708 ATGTGAAGTGAATGAGAAATGCACTGTCAATATGACTAG 2749


```

RESULT      4
AK024970
LOCUS       AK024970      2022 bp      mRNA      PRI      29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLN21317 fls, clone COL02275, highly similar to
              AP127036 Homo sapiens calcium-activated chloride channel protein 1
              (CACCI) mRNA.
ACCESSION   AK024970
VERSION     AK024970.1 GI:10437397
KEYWORDS    oligo capping; fls (full insert sequence).
SOURCE      Homo sapiens colon cDNA to mRNA, clone_1lb:COL clone:COL02275.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (sites)
AUTHORS     Kawabata,A., Hiki,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
            NEDO human cDNA sequencing project
            Unpublished (2000)
            2 (bases 1 to 2022)
            Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
            Direct Submission
            Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumio
            Sugano, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Structure Analysis, Human Genome Center;
            Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction: 5'-6' and one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES
SOURCE      1..2022
            Location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="COL02275"
            /clone_1lb="COL"
            /issue_type="colon"
            /note="cloning vector pME18FL3"
            1..2022
            /note="highly similar to AP127036 Homo sapiens
            calcium-activated chloride channel protein 1 (CACCI) mRNA"
BASE COUNT  612 a      472 c      453 g      485 t
ORIGIN
Query Match      40.4%; Score 1847.2; DB 9; Length 2022;
Best Local Similarity 95.2%; Pred. No. 0;
Matches 1955; Conservative 0; Mismatches 3; Indels 96; Gaps 1;

```

```

Db 241 GGGTTGGATGGTGCATTGTGACAGTGTGCTGCCCATGTACAAAGTAACTCATACAGATAA 300
Oy 2816 acaatgagcattgacagaggaacacactgcgcaaaagatacctgcagcagctccagaaggga 2875
Db 301 ACAGTGGCAGTGCACAGGAGACACACTCGCCAAAAGATACCTGCACAGCAGCTTCAGAGGGA 360
Oy 2876 cgtccatctgacaggggcttgatcgatctacgatataatgagcaaatctgctgttt 2935
Db 361 GGTTCATCTGCAGCGGGCTTCGATCGCATTTACT----- 395
Oy 2936 localgacacacagagattatgagagatgcgacaagaataatccaatctggcctctcgg 2995
Db 396 ----- 395
Oy 2996 cctgcagcttgatgattaggaagaatatccaactgatgatctgaatctgctgtcga 3055
Db 396 -----GTGATTTAGGAAGAAATATCCAACTGATGATGTGAATTTGCTGTGA 444
Oy 3056 cggatggggaagacacactataagtggtgctttaacgaggtcaacaagtggtgca 3115
Db 445 CGGATGGGGAAGACACACTATTAAGTGGGTCTTTAAACGAGGTCAAAACAAAGTGGTGCA 504
Oy 3116 tcatccacacagctgccttggtggccctctgcagctcaagaactagaagagctgtccaaa 3175
Db 505 TCATCCACACAGTGCCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGAGCTGTCCAAA 564
Oy 3176 tgacagagagtttacagacatatgcttcaagatcaagttcagaacaaatggcctcatatg 3235
Db 565 TGACAGGAGGTTTACGACATATGCTTCAGATCAAGTTCAGAACAAATGGCCATTTGATG 624
Oy 3236 ctttgaggcccttcacacagaatagagctgtctccagagccctccagcttaga 3295
Db 625 CTTTGGGCCCCCTTTCATCAGGAAGAGAGCTGTCTCAGGCTCCATCAGCTTGAGA 684
Oy 3296 gtaaggaatlaacctccagaacacagccagttgatlaagagcagttgctgtgacaga 3355
Db 685 GTAAGGATTTAACCTCCAGAAACAGCCAGTGGATGAATGGCAGCTGATTCGTGCAGAGA 744
Oy 3356 ccgtgggaaggaacacttgtttcttcaactcagtgacaatgcagctccccaatcttc 3415
Db 745 CCGTGGGAAGGACACTTGTGTTTATTCACCTGGACACAGCCAGCTCCCAATCCCTTTC 804
Oy 3416 tctggatcccaatgagaagaagcaggtgctgtgtgtagtggcaaaaaccacaaatgg 3475
Db 805 TCTGGATCCCAAGTGCAGAGCAAGAGTGGCTTGTGTGTGAGCAAAAACACCAAAATGG 864
Oy 3476 cctacctcacaatcccaaggcattgctlaaggttgacgttggaatatagctctgcaagaa 3535
Db 865 CCTACTCCAAATCCCAAGCATTGCTAAGGTGGCAGCTTGGAAATACAGTTCGACAGCA 924
Oy 3536 gctcaacaacttgacccctgactgtcagctccgttgcgtccaaatgctactcctgctcaa 3595
Db 925 GCTCAAAACCTTGACCTTGACTGTACGTCCGCGCTCCCAATCTACCCCTGCCCTCAA 984
Oy 3596 ttacagtgacttccaaaacgaacgaaggaacacagaataatcccgagccctcgtgagttt 3655
Db 985 TTACAGTACTTCCAAACGAACGAAGACACCAAGAAATCCCAAGCCCTCGAGTATT 1044
Oy 3656 atgcaaatatccgcaagagagcctcccaatctcaaggccagttgcaagccctgattg 3715
Db 1045 ATGCAAAATATTCGCCAAGAGAGCTCCCAATTTCTAGGGCCAGTGTCAAGCCCTGATTG 1104
Oy 3716 aatcagtgataggaanaaacagttaccttggaactatgataatgagcaggtgctgattg 3775
Db 1105 AATCATGTAATGGAAGAAACAGTTTACTTGGAACACTGTGATAATAGACAGTGTGATG 1164
Oy 3776 ctactaagatgagcgtgtctactcaaggtatttcaacaattatgacagaatgtaaat 3835
Db 1165 CTACTAAGATGACGCTGTCTACTCAAGGTATTTTACAACTTATGACAGAAATGTTAGAT 1224
Oy 3836 aacgtgtaaaagctgcggctctgaggaagatlaagcagcaggaagagatgatacccc 3895

```

DB	1225	ACAGTGTAAAGGCGCGGCTCTGGGAGGAGTTAAGCAGCCAGAGAGATGATACCCC	1284
OY	3896	agcagaagtgaagactgtatcataccgtgcgttgattgaagaatgaatgaatgaatgaatgc	3955
Db	1285	ACGAGGTGAGACGTGATCATACCTGGCTGGATTGAGATGATGAAATACATGTGAATC	1344
OY	3956	caccaagacctgaattaaagaatgaatgttcaacacagaagaagtgtttcaagcaga	4015
Db	1345	CACCAAGACCTGAATTATTAAGAGATGATTTCAACACAGCAAGTGTGTTTCAGCAGAA	1404
OY	4016	cattctcggaaggtcatttgttgcttcgtatgtccaaatgctcccaatacctgattctc	4075
Db	1405	CATCTCGGAGAGCTCATTTGTGGCTTCTGATGTCCCAATATGCTCCCAATCCTGATCTCT	1464
OY	4076	tcccaacttgcgaataaccagcagcttgaagaagggaattcaagggggagatccatataac	4135
Db	1465	TCCCACTGCGCCAAATACCCGACTGTAAGCGGGAATTTACGGGGCACTTCATTATAATC	1524
OY	4136	tgaacttgaacagctccctggggaatgaatgaacatgaagcaagctcaaatatcatc	4195
Db	1525	TGACTTGGACACTCTCTGGGAGATGATTAAGACCATGGAACACTCTCAAGTATATCATTC	1584
OY	4196	gaataagtaacaaagtattcttgatctcaagagacaagttaaatgaatctctcaagugaata	4255
Db	1585	GATTAAGTACAAAGTATCTTGATCTGATCTGAGAGACAAAGTCAATGAATCTTCAAGTGAATA	1644
OY	4256	ctcagctgtctcatcccaagaagaagccaactcgaagaagtcttttgtttaaccgaa	4315
Db	1645	CTACCGCTCTCATCCCAAGGAAGCAACTGTGAGGAATCTTTTGTTTAACCAGAA	1704
OY	4316	acattactttgaaatgagacagatctttcaatgtcatccaagctgttgataagctcg	4375
Db	1705	ACATTACTTTTAAATAATGCGACAGATCTTTTCATTGTCTTTCAGCGCTGTGATTAAGGCTCG	1764
OY	4376	acctgaatcagaataatcccaactgcagcatgatactttgtttatctccacagactc	4435
Db	1765	ATCTGAATACAAATATTCACACATTGACGAGATCTTTGTGTTTATCTCTCCACAGACTC	1824
OY	4436	cgcagagaacactatgtctgtatgaagaagctgtctctctgttccataatcatatcaaca	4495
Db	1825	CGCCAGAGACACCTGTGCTCTGATGGAAGCTGCTCTCTTCTTAATATTCATATCAACA	1884
OY	4496	gaacattcccttgagcatctccattttaaataattatgtgaagaatgagatagaagactgcagc	4555
Db	1885	GCACCATTCCTTGCCATTTCACATTTTAAATAATTATGTGGAAGTGAATGAGGAATGCGACG	1944
OY	4556	tgtaaatagacctag 4569	
Db	1945	TGTCATATAGCCTTAG 1958	
RESULT 5			
AF095584	3079 bp	mRNA	MAM 13-OCT-2000
LOCUS	AF095584	Sus scrofa epithelial chloride channel protein (ABCC) mRNA,	
DEFINITION		complete cds.	
ACCESSION	AF095584		
VERSION	AF095584.1	GI:6002645	
KEYWORDS			
SOURCE			
ORGANISM			
	Sus scrofa		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
REFERENCE			
AUTHORS	Gaspar,K.J., Racette,K.J., Gordon,J.R., Loewen,M.E. and Forsyth,G.W.		
TITLE	Cloning a chloride conductance mediator from the apical membrane of porcine ileal enterocytes		
JOURNAL	Physiol. Genomics (Online) 3 (2), 101-111 (2000)		
MEDLINE	20473747		
PUBMED	11015605		
REFERENCE	2 (bases 1 to 3079)		
AUTHORS	Gaspar,K.J., Gabriel,S.E., Racette,K.J. and Forsyth,G.W.		

FEATURES	source	Location/Qualifiers
gene	CDS	1..3079 /organism="Sus scrofa" /db_xref="taxon:9823" /tissue_type="ileal mucosa" 1..3079 /gene="AECB" 133..2886 /gene="AECB" /function="involved in exocrine chloride conductance" /codon_start=1 /product="epithelial chloride channel protein" /protein_id="AA00077.1" /db_xref="GI:6002646" /translation="MGSEFSLFLVHLLEGAQSNLSLIQNGNGEGLVIAIDPNP EDELILIKDMYTKASPLFEATREFYKNVALILPAMKAKPEVYRKLEKNKA DVVTEPENPNDPDEPYTEOMNGEKEKEKLYFPMDPAKGLVLOYGPGVFEHNAH LRKCFVENVYNEOKFVYLSNKKKPOPISSAIRGNVLPCCGSSCYRKPCRAADRYL FOKCEPIIPDPQSEKASIMPAQSIDTVVEFCCKHNKNCAPADONQKRNSTVEYL QDSDFKFTPTPTTPPAPEFSLQIQDIRVCLVDKSGMTVGRGLKRLNQKLEFL LQYEOGAWGMAWAFDSAAVYKSELVYINSAAERDALRSILPTAASGTVSICGLRSA FTYVKKRYPTDSEIVLTDGEDNTISACEPYKONALIHVALGSAKLELELSQ MSTGLOIYASPDQANNGLLIDAFGLSSGNSAORSIOLESOGITLIONNMMNGTVY DNGKDTPLILTELEKFLSPIPPEVPGSGRSDSLYGNHNKNAVPOVGTAKVGM KYSIQASSTITLTVSSRRSRTILPYTVYSKMKPDGKPPSPVYVYTRKHQSTPL RAKVALILESENGKTVTLELDNGAGDARKNDGITSRTFATIDANCERSVKWALGG VNTERRAPLEMSGAMIRGWIENGELIKMNPBRINDKLDQSCFSPATSGS ASDVKSPIDLPPEPKITDLKAGIQEDNLIINLTWAPGDYDGHGRADRYILISTNI LDLCKFNDSEVOVNTDILIRKANESEVPEVKEGELFTGTGDLFLAVQAVDNTKNS EINSIAQVSLFLPEADPEPPEPAPSLPCPEIQVNSTIPGIHLIKIMKMGELQL SLA"
BASE COUNT	892 a 758 c 755 g 674 t	
ORIGIN		
Query Match	35.8%; Score 1637.8; DB 4; Length 3079;	
Best Local Similarity	77.4%; Pred. No. 0;	
Matches 2084; Conservative	0; Mismatches 502; Indels 105; Gaps 4;	
1888	aagacatggtgacccagagcatctctgatactgcttgaagctacagaagaagcttcat	1947
292	AAGGATATGGTGACCAAGCATCTCCATATCTGTTGAAGCTACGAGAAAAGATTCAT	351
1948	ttcaaaaatgttcgacatttgaatccctgaacaacatggaagaacaagctgactatgaga	2007
352	TTCAAAATGTGGCATTTTATGTTCTCTCAAGTTGGAAAGCAACCAGATATGTGAA	411
2008	ccaaaacttgagaccatacaaaaatgctgaatgcttgcgtgttcgtagctactactccag	2067
412	CCCAAACTTGAGACCTACAAAAATGCTGATGTTGTGGTTACTGAAACCAAAATCC	471
2068	aatgataaaccttcacacagacagatggcagcagtgtagaagaaggttgaaagaatcac	2127
472	AATGATGGGCCCTATACGGAGCAATGGGAAACGTGACAGAAAGGCTGAAAAATTTAT	531
2128	ctcaactctgatttcaatgacgagaaaaaagttaactgtaacttgagacaacaagtaagca	2187
532	TTCACTCTGTGACTTTCGTATGACGAGAAAAAGGTGCTTCATATGACCAACAGAAAGG	591
2188	tttgtcaatgagtggtcatalctacgatggggaglatlttgacgagatacaataatgatgag	2247
592	TTTGTCCAGCATGGGCTCACCTCGATGGGGCGCTGTTTATGATGATCAACACAGACAG	651
2248	aaattcactatcaacaatgagaagaatacaagcagtaagatgattcaacagttatcagct	2307
652	AAGTTCTATTATCCACAAAAAGAACACCACTGATATGTTTGGGAGCCATCCGGGT	711
2308	acaatgtagtaagaaggtgtcagagagagcagctgttacccaacaaagatgcacattcat	2367

Dh 712 ACAATGTACTACCAAAATGCCAAGGAGGACGTGTGACCAAAACGTGCCGAGCAGAC 771
Qy 2368 aagaataacagagctatgaataaagatgtagttgtctccaaatccgcgcagacgag 2427
Dh 772 AGATTAAGGGGCTGTCCAGAGAGAGTGTAGTTCAATCCCGATCCCGACAGAGCGAG 831
Qy 2428 aaggtctataatgattgtgcacaacatglttatctatagttgaattctgtacagaaca 2487
Dh 832 AAGCCTCTAATATGTTGACAGAGCATGTGATCTGTGTTGAATCTGTATAAGAAAA 891
Qy 2488 aaccacaacaagaagctcccaacaagaanaatacaaaatgcaatctccgaagacatg 2547
Dh 892 AACCAATATAAGAAAGCTCCAAAGCAACCAAAAAATCAATCTCGAAGCAATGG 951
Qy 2548 gaagtacgcgtgattctggaagacttaagaanaacctctatagacaacaagacacca 2607
Dh 952 GAATGATCCAAAGTTTGGAAGACTTTAAGAAAACACTCCCATGACAAACAGCACACC 1011
Qy 2608 aatcccaactctcatatgctgcagattggacaagaattgtgtttagtccctgacaa 2667
Dh 1012 GCGCCACCTTCTCACTGCTGACAGTCCGACAGAAATCGTGTCTAGTCTCGACAA 1071
Qy 2668 tctggaagcatgagcgactggtgaacgcctcatatgcagatcaagcgccagctttc 2727
Dh 1072 TCTGGAACCATGACGGTTGTGCGCTTAAAGCATGAATCAAGCAGGAAACTCTTC 1131
Qy 2728 ctgctgcagacagttgagctggtgctggtgttgagatggtgacattgagacagttgctc 2787
Dh 1132 CTTCGACAGACAGTGGACCAAGGGGCTGGGTGGGATGGTGACATTGACACGCTGCC 1191
Qy 2788 catgtacaanaatgaaactatacaagaataacagtgagtgacagaggaacacatgcacaa 2847
Dh 1192 TATGTAAAAAGTGAACCTGCTCACATAAACGCGCGGAAAGAGAGCGGCTGCCGC 1251
Qy 2848 agattactctgacagagctctcagagagcgctccatctgcaagcggtctcgatcgcatct 2907
Dh 1252 AGCTTACCACAGCAGCTCGGAGGAGAACGTTCATCTCGTGGGCTTCGCTCCCATTT 1311
Qy 2908 actgatatgtgcaaatcttgctgttttcatgtacacagcagctttagtggagtgga 2967
Dh 1312 ACT----- 1314
Qy 2968 caagaaatccaatltggcctctctgctgcagcttagtgaatagaagaatatcca 3027
Dh 1315 -----GTGATCAACGAAGAAATACCG 1335
Qy 3028 actgatgatctgaatgtgtctgtgacgagatggaggaagaacacataagatgagtc 3087
Dh 1336 ACCGACGGGTGTGATGTGTACTGACGACGGGAGGACAAACGATAAAGCGCGTGC 1395
Qy 3088 ttaacgaggtcaacaagaatggtgctcatccacacagtcgctttgggcccctctga 3147
Dh 1396 TTTCGGAGGTGAAGCAAAAGGAGCCATCATCCACAGGTGCGCCTTGCGGCCCTCGGCA 1455
Qy 3148 gctcaagaactagagagctgttccaanaatgaacagaggtttacagacatagctcagat 3207
Dh 1456 GCGAAGGAGCTGGAGAGCTGTCCAGATGACAGGTGTTGACACATATGCTTCAGAC 1515
Qy 3208 caaattcagaacaatgagctatgattgtctttgggccccttcaatcaagaatgagat 3267
Dh 1516 CAGCCCGAACAACGGGCTTATTGACGCTTCGGGCGCTTTCCTCGGGGAACGACCT 1575
Qy 3268 gctctcagcgctccatccagcttgaagatgaagatcaacccctcaagaacagcagtg 3327
Dh 1576 GCGTCCGACGCTCATCTACACTGAGAGTCAAGGCTTAACCTTCACAAACAGCAATGG 1635
Qy 3328 atgaatgacagatgattcgtgagacacagctgggaaagaacacttgtttctatcac 3387
Dh 1636 ATGATGCGACAGTGTGTGGAGACGACTGTGGAAAGGACACTTGTCTCATCAC 1695
Qy 3388 tgaacaaatgag-----cctcccaaatctctctc--gggatccagtgagcagaagaa 3441
Dh 1696 TTGGAACGGAAGTCTTTCCTCCCAATTCCTTTTGGGTTCCCAAGGAGGAGCCAG 1755

Qy 3442 gttgctttagtggacaaaacacccaatggcctacccccaatcccaagctatgc 3501
Dh 1756 GACACTTTCCTTGTTGGAANAACACAAAGATGCTTACTTCCAAAGTCCAGGACGGCC 1815
Qy 3502 aaggttgacacttggaaatacagctctgaagaagctcacaaaccttgaacctgactgc 3561
Dh 1816 AAGGTGGCATGTGAATACAGCTTCGACGAAGTTCAAAACCTTCATTTGACCGTC 1875
Qy 3562 acgtccgttggtccaatgctacccctgctccaattgaagtgacttccaagaagaag 3621
Dh 1876 AGTCCCGTCCGTCAGGCTACCTGCTCGGTTTACCTGTGACCTCTAAATGAACAA 1935
Qy 3622 gacacacaatcccccagccctctgtagtttaatgaatatccgaagaagaccc 3681
Dh 1936 GACACGGCAATTCGCCAGCCCATGCTAGTTTACAGGAATTCACCAAGAACCTCG 1995
Qy 3682 ccaattcgaagccagtgacacagccctgattgaatcagtgaaatgaaacagttacc 3741
Dh 1996 CCCATTCTCAGGGCCCAAGGTACACACCTCGATAGCAATGAGAAATGGAAAAACATTACC 2055
Qy 3742 ttggaactactgataatgagagcagtgctgattgattactaagatgaagtggtctacca 3801
Dh 2056 TTGGAATTACTGGACAAAGGACAGCGCTGATGCTTACCAAGAAATGAGGATGTACTCG 2115
Qy 3802 aggtattcacaacttgaacaagatgtagatacagtgtaaagtgagcgcttgga 3861
Dh 2116 AGGTATTTTACAGCTTACGATGCAAAATGTGATACACGTRAAAGTGTGGGCTTGGA 2175
Qy 3862 ggaagtaacgagcagcagagagtgatacccccagagatggaacatgatacact 3921
Dh 2176 GGAGTGAACACGCGCACAGAGGACCCCTCTGAGAGGAGCCATGATACGT 2235
Qy 3922 ggtctgatttgaagatgaataataaataagaaatcccaagaacctgaataataagat 3981
Dh 2236 GCTGATTTGAAGTGGGAGATCAAAATGGAACCGCGGACCGCAATTAAACAGAGAC 2295
Qy 3982 gatgtcaacaacaagatgtgttcaagcagaacatctcgagagagctatcttggct 4041
Dh 2296 GACCTTCAAGGGCAAGATGTGTTCAAGCCGACAGCTCGGGGGTTCATTGCTGGCC 2355
Qy 4042 tctgaltcccaaatgctcccaatcctcctccacactgagcaaatcaacagcactg 4101
Dh 2356 TCCGAGTTCGGAAGTCTCCCATACCTCATCACTGAGGTGAGCGCTCTGGGATGAT 2415
Qy 4102 aagcggaatctcaggggagagctcatttaactcgtacttgagacgctccgggagat 4161
Dh 2416 AAGCGGGGATCCAAAGGGACACTCATCACTGAGGTGAGCGCTCTGGGATGAT 2475
Qy 4162 tatgaccaatgaacagctcaacaagatataatctgaatgaatgaatgaatctatctc 4221
Dh 2476 TACGACCAAGGAGAGTCAAGAGTACATCATCAATTAAGACAAATATTTCTTATCTC 2535
Qy 4222 agagaagaatccaatgaatcctcctaagtgaatactactgctcatalcccaagaagcc 4281
Dh 2536 AGAGCAAGATTCAATGATTTCTGTTCAAGTGAACCAAGATCTCATCCAAAGAGGCC 2595
Qy 4282 aactctgagagatcttltgttaaccagaacaatcattactttgaaataggcagaat 4341
Dh 2596 AACTGAGAGAGGTCTTGTGTTAAACCAAGGACATCCCTTTTCAAAATGGCAGAT 2655
Qy 4342 ctlttcatctatccagcgtgtatgaatgagtgatgcaatcagaatatccaaact 4401
Dh 2656 CTCTTCAATTGCTGTAAGAGCTGTGCATTAAGACCAACCTGMAAGTCAAGAAATCTCCAACTT 2715
Qy 4402 gaaagatcattgttattctctccacagactccgcagagaacact--agtcctgat 4458
Dh 2716 GCACAAAGTATCTTGTCTTCCCGGAGAGCTCTCGGAGACACTTCGAGAGACACT 2775
Qy 4459 gaaagctctgctctctgctcctaataatataatcaacagcacatctcgtgattcaact 4518
Dh 2776 GCTCTCTCTGCGCTTGTCTGAATTAAGTCAACACACCATTCCTGAGCAATTCACATTT 2835

OY 4519 ttaaaatttctggaagtagataggaagcagctgtcacaagcctag 4569
|||||
Db 2836 TTTAAAAATTTATGTGAGTGGCTGGAGAAATACGCTATCGATAGCCTAG 2886

RESULT 6
AB017156
LOCUS AB017156 2937 bp mRNA ROD 10-NOV-1999
DEFINITION Mus musculus gob-5 mRNA, complete cds.
ACCESSION AB017156
VERSION AB017156.1 GI:3721911
KEYWORDS GOB-5.
SOURCE Mus musculus adult intestine goblet cell CDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Komiya,T., Tanigawa,Y. and Hirohashi,S.
TITLE 1 (sites)
JOURNAL Cloning and identification of the gene gob-5, which is expressed in
MEDLINE intestinal goblet cells in mice
REFERENCE Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)
AUTHORS 2 (bases 1 to 2937)
JOURNAL Komiya,T., Tanigawa,Y. and Hirohashi,S.
TITLE Direct Submission
REFERENCE Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru
AUTHORS Komiya, T., Tanigawa, Y., Genome Asymmetry Group, Doi Biosymmetry
JOURNAL Project, Tokodai 5-9-9, Tsukuba, Ibaragi 300-2635, Japan
TITLE (E-mail:tkom@bio.erauto.trc-net.co.jp, Tel:81-298-48-1515, Fax:81-298-47-8901)

FEATURES
source location/Qualifiers
1. 2937
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="goblet cell"
/dev_stage="adult"
/tissue_type="intestine"
15..2756
/gene="gob-5"
15..2756
/gene="gob-5"
/codon_start=1
/protein_id="BA33743.1"
/db_xref="GI:3721912"
/translation="MESLSPVFLILHLEGLVLSLSLIQLNNNGYEGIVIAIDHDP
EDEALIOHKIDMTQAPYLFATGKRFYKFNVAIIPSMKAPPTPKTETENA
DVLVSTPLGNDPEYTHIGAGEGKIRHLIPDLACKLKLQYRPOORTFVHEVAH
FRMGVENEYNNDEKEFLSKGKPOAVGSAITGKNVRRCCGGSCITTKCVDRVYG
LYKNCVFPYPPDHONKASIMRNQINSYVERTEKNNHQBEPNDONORCNLRSTVEY
IOESDEKQTPPTAPPAFTSLDIOGRIVCLVLDKSGSLINDRLNRMNQSLRF
LQIYEOGSMGVMYTFDSAAVYVOSLQNSGARDLILKHLPTVSAGGTSICSGLRT
AFTVILKKYPTDGSSEIVLLTDEGNTISCFDILKOSGAIHTHVALGPAARAELEDS
KMTGLOIYSSDOVONNGLVDAFAALSGNAALIAHOSIOLESQVNLONNONGSVI
VSSVSGKDLFLITHTHTPTFTFIMPSDEONGFTLIDTTRVAYLOVGTCAKVGWK
YSLQASSQTLTLVTSRASATLPTTPYVYVKNKTKGKPPPYVYASIRQASPLR
ASYVALIESVNGKTVYLELLDAGADATKNDGVSRFTAPDANRISVKIKALGVG
TSDROAPKPKNRAMIDSMIDGDEVNMPPRETSYVODKQLCFSTSSGFEVATN
VPAAPIPDLPPPCQITDLKASIOGONLVNLWTAPAGDDYDHGRASNTIIRSTSTVD
LBDHFTSLQVNTGLIPKEASSEIEFEELGNGTPGCDIFIAIOAVDKSLKSEI
SNIRSVYIIPNOEPIPEDSTPCPDISINSTIPIQIHVILKIMMKLGEQVTLGH"

BASE COUNT 860 a 718 c 693 g 666 t
ORIGIN

Query Match 33.2%; Score 1517.6; DB 10; Length 2937;
Best Local Similarity 75.0%; Pred. No. 0;
Matches 2017; Conservative 0; Mismatches 554; Indels 117; Gaps 5;

OY 1888 aaggaatcgtgaccagcagcattctgtatctctgaagctacaggaagcagattat 1947
|||||
Db 174 AAGGACATGCTGACTCAGCGCTCTCCATACCTTTTGAACCTACAGAAAMAAGATTTTAC 233

OY 1948 ttcaaaaatgtctgcatcttgattcctgtaaacatggaagacaagcgtgactatgtgaga 2007

Db 234 TTTCAAAATGTTGGCATTGTTGATTTCCGAGAGCTGGAGGCAAGCAAGCCTGATATACAGG 293
OY 2008 ccaaaacttgagacctacaaaatgtcgtatgtctgtgctagtactccccaagt 2067
|||||
Db 294 CCAAACTTGAAACCTTCAAAAACGCTGATGTCTTGTATACACACACACCTCTTGAGC 353
OY 2068 aatgatgaaccctacactgacagatgaggaactgtgagagaagagtgaaagatccac 2127
|||||
Db 354 AATGATGAGCCCTTACACCGAATATATAGAGCATGTGAGAAAAAGGGATACAGATTAC 413
OY 2128 ctcaactcctgatttcatcttgcaagaaaagtgtgcgtatataatgaccacaaagttaggca 2187
|||||
Db 414 CTGACTCTCTACTTCTTGAAGGAAAGACCTGACTGATAGTGGCCACAGACAGAGACC 473
OY 2188 ttgtccatgagtgagctcatctacagatgggagatatttgacagatcaataatgagag 2247
|||||
Db 474 TTTGTCCATGATGGCTGCTCATCTTCGATGGGAGCTGTTATGATATCAACAAACGACAG 533
OY 2248 aaattcacttaccatggaagaatacaagcagtaagatgttcagcaggtattactgtgt 2307
|||||
Db 534 AAGTTCTACTTATCCAAAGCAAAACCCCAAGCAGTGAAGTTCAGACGCTATCCGGT 593
OY 2308 acaatgtatgaagaagtgtcagggagcagctgttaccacca---aagatgcacattc 2364
|||||
Db 594 AAAATTCAGTTTCGTGGTCCGAGGAGGAGCAAGTTGTACTAAGGAAAGTGTATATC 653
OY 2365 aataaagtaacagactatgaaaagagatgtgagattgttcccaatcccgacagag 2424
|||||
Db 654 GACAGAGTAAAGGAGCGTGTATTAAGCAATTTGTATTTTGTACCAAGTCCACACCAAAAC 713
OY 2425 gagaaggtctctataatgttctgcaacaactgttgattctataagttgaattctgtacaga 2484
|||||
Db 714 GACAAGGCTTCCATCATGATTAAACCAAAATATCAATCTGTGTTGAATTCGTACAGAA 773
OY 2485 caaaaccacacaaagaagctccaaacagaacaaatcaaaaatgcaatcccggaagaca 2544
|||||
Db 774 AAAAATTCACATCAAGCAAGGCCCAAAAGCAAAACCAAGATGCAATTCGCAAGACG 833
OY 2545 tgggaagtgtacgtgtgtctgagagacttaagaaaaccctctctgacaacacagcca 2604
|||||
Db 834 TGGGAATCATTCAGAGATTTGAGAGCTTACAGCAAAACCTCCATGACAGCCACCA 893
OY 2605 ccaatccacacttccatctgctgacagattggaacaaagaattgtgtgttagtccctgac 2664
|||||
Db 894 CCGACACCCACTTTCATCTGCTGCAAAATGTGACAAAGAAATTTGTGCTTATGTTGAT 953
OY 2665 aaatcgtgaagcatgtgcgactgtgtaacgcctcaatcgcactgataagaagcagcact 2724
|||||
Db 954 AAGTCCGGGAGCATGCTGAACGATGATCTTAAACCGAATGATACAGCAACCGGCTT 1013
OY 2725 ttccctcgtgagaagaattgagctgggggtccctgggttgggaatgtgacatttgcagtgtc 2784
|||||
Db 1014 TTCTCGTGTGAGACTGTGAGAGGAGGATCTCGGTGGGATGAGTGTACCTTTGACATGCT 1073
OY 2785 gcccatgtacaaaatgaactacacagataaacagatgtgcagtgacagggagacacactggcc 2844
|||||
Db 1074 GCTATGATCAAAAGCAACCAACCAAGATTAAACAGTGTGCTGACAGAGATCTGTGATG 1133
OY 2845 aaaaagattaccctgacagcagcttcaaggaaggaagcttcaatcgtcagcggggtctgacatggca 2904
|||||
Db 1134 AAGCAGCTTACCCACAGTATCTGAGGAGGAGACATCTATATGCTCTGTGGCTTCGGAGCA 1193
OY 2905 tttaactgatatgtgycacaactttgcctgttttccatgacacacagcagattatggggagt 2964
|||||
Db 1194 TTTTAC----- 1198
OY 2965 cgaacaagaatccaaatgtggcctctctgctgacgttgaatgtatgaggaagaatat 3024
|||||
Db 1199 -----ATTGATTAAGAAGATGAT 1217
OY 3025 ccaactgtatgagatgtgaaatgtgctgtcgtgacgagatgtgggaagaacaacacataatgtgg 3084
|||||

Db 1218 CCAACTGATGATCTGAATTGCTGCTGACCGATGGGGAGACAAACCAATTAGCAGC 1277
 Qy 3085 tgccttaacagagtcacaagaatggtgccaatccacacagtcgcttggggccct 3144
 Db 1278 TGCTTTGACCTGGTGAAGACAGGGGGCCATCATCCATACAGTGGCCCTGGGACCGCT 1337
 Qy 3145 gcaactaagaactagaggagctgtccaaaatgacagaggtttacagacatgcttca 3204
 Db 1338 GCCCTAAAGAGCTTGAGACACTGTCCAAATGACAGGAGCGCTGCAGACATACCTTTCG 1397
 Qy 3205 gatcaatgacgaacaatggcctcatgtagctgttgggccccttcaatgaagaaatgga 3264
 Db 1398 GATAGGTTGGAACAAATGGCTTGTGATGCTTTCGACACTCTCCCTGGAATTCG 1457
 Qy 3265 gctctctcactagcgtccatccagcttgagtagtaaggatcaacctccaaaagagcag 3324
 Db 1458 GCGATCCGTCAGCACTCCATCCAGCTGAGAGACAGGGAGATTAACTCTCAAAATTAACCA 1517
 Qy 3325 tgaatgaatgacagatgatalcgtgacagacgctgggaaaggaactgttcttca 3384
 Db 1518 TGAATGATGGCTCAGTATGATGAGACAGCTGGGGAAGACACCTGTTCTTATTC 1577
 Qy 3385 accctggacaatgacagcctcccaaatcctctcgggataccagtgacagaagaagct 3444
 Db 1578 ACCCTGGACAAAGCATCCCTCTACAAATATTTATCTGGGATCCAGGAGTGAACAAAT 1637
 Qy 3445 gacctttagtgagcaaaaacacaaaatggcctacactcaaatcccaagatgcttag 3504
 Db 1638 GGTTTTATCTAGACACACACACTAAGTGGCTTACCTCCAACTCCAGGACGGCTAG 1697
 Qy 3505 gttggcacttggaaatcagctctgcaagcaagctacaacacttgaacctgactgacg 3564
 Db 1698 GTGGCTTTTGAATACAGATTCACAGCGAGCTCACAGATCCTTACCTTACCTGAC 1757
 Qy 3565 tcccggtgctcaatgctaccctgcctcaatitacagtgacttccaaaagaaagagac 3624
 Db 1758 TCCCGTGACGAAGTGTCTACACTGCTCTCTATCTACAGTACCGCGTGTGATTAAGAAC 1817
 Qy 3625 accagaatattcccaagccctgtgtagttatcaaatatgcccagaagagccctccca 3684
 Db 1818 ACAGGGAATTTCCCGAGCTGTATACAGTGTATGACACATTCGCCAAGGAGCCTCGCT 1877
 Qy 3685 atctcaggggacagtgctcaagccctgaltgaatcagtgaaatgggaaagaaatcagctg 3744
 Db 1878 ATCTCAGGGCCAGCGTACAGCCTTGATGATCTGTGATGGAATAAAGTAACTACCTG 1937
 Qy 3745 gaactactgataatgagagagtgctgtagtactaagaatgacagtgctactcaag 3804
 Db 1938 GAATTACTGGATTAACGGAGCGGTGCCATGCCACCAAGAAATGATGTTCTACTCAAG 1997
 Qy 3805 tatttcaactttagacagaatgtagatagacagtgtaaaatgggggctctgggaaga 3864
 Db 1998 TTTTTCACAGCTTTTGAATGCAATATGATGATACAGCTTAAATATGGGCTCTGGGAGGA 2057
 Qy 3865 gttacagcagcagagagagtgatatacccaagcagatggagacatgatactactg 3924
 Db 2058 GTCACTTCAGACAGACAGAGAGCAGCAGCTCCGAAGAAGACCCATGTACATATAGGC 2117
 Qy 3925 tggatggaatgataatcaatcaatgatacccaagacactgaaatgaagaatgag 3984
 Db 2118 TGGATTTGAGGTGGTGAATGAATGAACCCACACCTCTCTGAAACATGTT-----AT 2171
 Qy 3985 gttcaacacagcaagatgtagtttcaagcagaacatcctcggagagcttctggtctc 4044
 Db 2172 GTTAAAGCAAGACAGCTGCTTCAAGAGACATTTTCAAGGGGATCTTTGTGGCCAC 2231
 Qy 4045 gatgtcc---caatgtctccataactatctctcccaactgycgaatataccagcagct 4101
 Db 2232 AATGTCCCGAGAGAGAGCTCCATTCCTGACCTTTCCACCTGTCAATATCAGTACCTG 2291
 Qy 4102 aagcgagaatctcagggggagctcattcaatctgacttgagacagctcctggtggaatg 4161
 Db 2292 AAGGCACAGATCCAAAGGAGAGAACTGGTGAATCTGAGTGGACGGCTCTGGGGATGAC 2351

Qy 4162 tatgaccatggaacagctcacagaatatactcgaatgaagaatcattcttgaatc 4221
 Db 2352 TACGACCAAGGAGAGCTCTTCACTACTCATTCGAAATGACACAGATGTGTTATCTC 2411
 Qy 4222 agagacaacttcaatgaaatctcttcaagtgaatactactgctctcaltcccaaggaagcc 4281
 Db 2412 AGGACCACTTCAACACTCTCACTCAAGTGAACACTACCGGCTTATCCCAAGAGGCC 2471
 Qy 4282 aactctgagaagctcttcttgaataaccagaataactacttcttgaataatggcagagat 4341
 Db 2472 AGCTTGAGGAATCTTGTGATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2531
 Qy 4342 ctcttactgctatcgaagcgttgaatgaagtgctgtaataatgaataatcaact 4401
 Db 2532 ATCTTCACTTCTATCCAGGCTGTGATGATGATGATGATGATGATGATGATGATGATG 2591
 Qy 4402 gcaagagatcttltgtatctctccacagactccgcagagacacactagctcgtatgaa 4461
 Db 2592 GCAGGGGTCTGTGTTTCATCCCGCTCAG-----GAGCCGCTATCCGAAAGAC 2642
 Qy 4462 acgtctgctcttctccaaatataatcaatcaacagacacactcctgcatcaattta 4521
 Db 2643 TCAACTCCCTTGTCTGATCATGACATCAACACACACATCTCTGATCCACGCTGCTG 2702
 Qy 4522 aaatattgtgaatgagatgagaagactgcagctgtcaatgacctag 4569
 Db 2703 AAGATATGTGGAATGGCTTGAAGGGAATGCAAGTGAACATGATGATGATGATGATGATG 2750

RESULT
 AF127035
 LOCUS
 DEFINITION
 mRNA, complete cds.
 ACCESSION
 AF127035
 VERSION
 AF127035.1 GI:5726288
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 3204)
 AUTHORS
 Agnel,M., Vermet,T. and Culouscou,J.M.
 TITLE
 Identification of three novel members of the calcium-dependent
 chloride channel (CaCC) family predominantly expressed in the
 digestive tract and trachea
 FEBS Lett. 455 (3), 295-301 (1999)
 JOURNAL
 MEDLINE
 PUBMED
 99364503
 10437792
 REFERENCE
 2 (bases 1 to 3204)
 AUTHORS
 Agnel,M. and Culouscou,J.-M.
 TITLE
 Direct Submision
 Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des
 Carrieres, Ruell-Malmaison 92500, France
 FEATURES
 source
 1..3204
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="colon"
 1..3204
 /gene="CaCC2"
 29..2782
 /gene="CaCC2"
 /note="bovine epithelial chloride channel homolog"
 /codon_start=1
 /product="calcium-activated chloride channel protein 2"
 /protein_id="AAD48398.1"
 /db_xref="GI:5726289"
 /translation="MGLEFRGVFLVLCILHQSSTSLKLNNGFEDIVIVYDPSVE
 DEKTIIEDVNTASTYLEATKRFKFNKSLILENNKFNQYVRPHENKRAHD
 VIVAPPLPGDEPYTKQFECGKGEVIEHTPDLLGKQNEVGPGLFVHMAHL
 RMGVFDEYEDQPYRAKSKRIEATRCASIGISGRNRYVKCGGSGLSLRACRIDSTKL

YSNGRCOFPPIVYOEKASIMFMSIDSVYECNECHTHNOEASLONICNGRSMWEYI
 XNBSDEHNTIDPMTPPPPEPPLSLTROHIVTILDKSGMGRNBRMNAOKHFL
 FOYJENSMWGMWHEPSTATVYKLOIKOSDOPRMTIAMPDPTVIGSTISGICIKYKA
 FOYGEHLSQDSSEVLLTDEEDNTASSCIDEVYOSCALVHEFALGNADEAIEMMS
 KRTGSHFHYVSEAKONGGLDARGALTSNTDLSQKSLQSLKSLGJLNSAMNDYI
 IDSVYVQDFFELTWNLSNPSISLMPSTIMENFTVATLSTSMALISLPGIARGVMA
 YNLOAKANPELITVYSTRANSVSPITVYNAKMKDVNSPSPVYAEILOOYVYV
 LGANVAFIESEONGHEVLLDNGEAGOSFNDVSYKSPFAYENRSLKARANGH
 GANVATRIKLEONHRAHYIGVWVNEIENANPREIDEDTLEDSFASGACVAY
 VSOYSPLEPOQVPSOITDLDATVHEKDLITLTAAGNENGVQVORYIRIASLIL
 DRUSPDALOVNTDLSPEAKSKSPAFKENCENISEENATILPAIKSDIKSLNUTSK
 VTSYAOYTLFIPAOANDPIDPPTPIPTPKDSKNSGVNITSLVLSIGSVYVNFILS
 VTSY


```

Db 1953 TTTTGGAACTTTTGGATAATGGTGCAGCGCTGATTCCTTCAAGAAATGATGAGCTACT 2012
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3800 caaggtatctccaaactatgacacgaaatgtaatacagtgtaaaagtgcgctcgg 3859
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2013 CCAGGTATTTTACCATATACAGAAATGGCAATATAGCTTAAAGTTGGGGCTCATG 2072
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3860 gaagagttaacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3919
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2073 GAGAGCAAAACACTGCCAGGCTAAATTTACGGCTCCACGTAAATAGAGCCCGCTACATAC 2132
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3920 ctggctgagatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg 3979
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2133 CAGGCTGGGTAGTGAACGGGGAATTTGAAGCAACCCGCCAAGACCTGGAATTTGAT---G 2189
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3980 atgagtctcaacaacaagaagcagcagcagcagcagcagcagcagcagcagcagcag 4039
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2190 AGGATACTCAGACACCTTTGGAGAGATTTACGCCAGACAGCATCCGAGGTGCTATTTGG 2249
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4040 ctctgagtctcccaaatgctcccaatgctctcccaatgctctcccaatgctcccaatg 4099
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2250 TATCACAAGTCCCAAGCTTCCCTTGGCTGACCAATACCCCAAGTCAAAATCACAGACC 2309
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4100 tgaagcgcgaatcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4159
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2210 TTGATGCCACAGTTCATGAGS---ATAAGATTATTTCTTACATGACACGACGACGAGATA 2366
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4160 attatgcaatgcaagaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4219
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2367 ATTTTGTGTTGGAAAGTTCACGTTATATATATATATATATATATATATATATATATAT 2426
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4220 tcaagaagaagtcgaatcctcgaatcctcgaatcctcgaatcctcgaatcctcgaatcctc 4279
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2427 TAAGAGACAGTTTGTGATGATGCTCTTCAAGTAAATACTAGATCTGTCACCAAGAGAG 2486
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4280 ccaactgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 4339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2487 CCAACTCCAGGAAGAAAGCTTTGATTTAAACCAAGAAATATCTCAGAAAGAAATGCAACCC 2546
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4340 atcttctcatctgcatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4399
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2547 ACAATTTTATTTGCTATTAAGTATTAAGTAAAGCAATTTGACATCAAAAGTATTCACAA 2606
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4400 ttgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 4459
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2607 TTGACACAGTAAGCTTTGTTTATTCCTCAAGCAATCCGATGATGATGATGATGATGAT 2666
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4460 aaacgtctgcctctgcctaataatcata 4489
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2667 CTCCTACTCCTACTCTGATTAAGTCATA 2696
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) to the DDBJ/EMBL/Genbank databases, Sumio
Sugano, Institute of Medical Science, University of Tokyo, Department
of Virology, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna1@elms.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by science and Technology
Agency).
FEATURES
     source             1..3221
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="COL01613"
                        /clone_1lb="COL"
                        /tissue_type="colon"
                        /note="cloning vector pME18SFL3"
     misc_feature       1..3221
                        /note="highly similar to ECLC_BOVIN EPITHELIAL CHLORIDE
                        CHANNEL PROTEIN"
BASE COUNT  1105 a      600 c      634 g      882 t
ORIGIN
Query Match 24.7%; Score 1129.8; DB 9; Length 3221;
Best Local Similarity 67.1%; Pred. No. 1.1e-279;
Matches 1749; Conservative 0; Mismatches 747; Indels 111; Gaps 6;
Qy 1889 aggcacatgtagccacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1948
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 200 AGGATATGTTGATACACCTTTCTACGTTACGTTGGAAGCCAGAAATATTTT 259
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1949 tcaaaaagtgtccatttgatcttgaaacatgaaagcagcagcagcagcagcagcagcagc 2008
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 260 TCAAAATGTATCTATTAATTTCTGAGATTGGAAGAAATTCCTCAGTACAAAGGC 319
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2009 caaaactgagacacacaaatgctgagctgctgctgctgctgctgctgctgctgctgctg 2068
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 320 CAAAACATGAACCAATTAACATCTGATGTTAGTTGACACACCTACCTCCAGGTA 379
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2069 atgagtgaacctacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2128
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 GAGATGAACCATACCAAGCAAGTTCACAGATGTGGAGAAAGCGAATATTCAC 439
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2129 tcaactcctgattcatcagcaagaaagctgagcagcagcagcagcagcagcagcagcagc 2188
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 440 TCACCCCTGACCTTCTACTTGGAAAAAACAAATGATATGACACACGCAACACTGT 499
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2189 ttgtccatgagtgagcctacatcagatgagcagcagcagcagcagcagcagcagcagcagc 2248
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 500 TTGTCCATGAGTGGGCTCACCTCCGGTGGGAGTGTTGATGAGTACAAATGAATCAGC 559
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2249 aatctctctatccaatggaagaatac---aagaacgaagaagtgctcagcagcagcagc 2305
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 560 CTTTCTACCGCTGTAAGTCAAAAAATTCGAAGCAACAGGTTCCGACAGTATCTCT 619
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2306 gtacaatgtatgaagaagtgctcagcagcagcagcagcagcagcagcagcagcagcagcagc 2365
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 620 GTAAATATAGAGTTTATATAGTGTCAAGAGGACCTGTCTTATGATGACATGCAAGATTG 679
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2366 ataaagtaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2425
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 680 ATTGTACAAACAAACCTGATGAAAGATGTCTTCTCGATTAAGTACAAACAG 739
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2426 agaagcctctataatgttgcaacaacatgctgctcagcagcagcagcagcagcagcagcagc 2485
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 740 AAAAAGCATCATATATGTTTATGCAAGATGATTTGTTGTTGGAATTTTGTACGAA 799
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2486 aaacccacaacaagaagcctccaacaagaagaatacaaaatgcaatctccgaagcagcagc 2545
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 800 AAACCCATTAATCAGAGCTCCAGCTACAAAACATTAAGTGAATTTTGAAGTACAT 859
QY 2546 gggagatgacccgtatctcgagactttaagaanaaccactcatgacacagagc 2605
Db 860 GGAAGGATGATTAACAAATTCGAGATTTTAAAAACACCATACCATGCTGACACACTC 919
QY 2606 caaatccacccttcacatgctcgagatgagcaaaagaattgltgtatgacctgaca 2665
Db 920 CTCACCTGCTTCTCTCATCTGCTGAAGATCCGTCAAAAGAAATGTGTGCTAGTTCTGATA 979
QY 2666 aatctggaacatagggacggtlaaccgctcactgaactgaatcaagcgccagctt 2725
Db 980 AGTCTGGAACCATGGGGGATGAAGACCGCTTAATCGAATGAATCAAGACGACAAACATT 1039
QY 2726 tccctgcgcagacgattgagcgtgggtcccggttggatggtgacatgacatgctg 2785
Db 1040 TCCTGCTGCAGACTGTTGAAATGATCCTGGGGGATGCTTCACTTTGATGATGACTG 1099
QY 2786 cccatgtacaanaatgaaatcacaagaataacagtgagcagtgacagggacacactgc 2845
Db 1100 CCACATATTGTAATTAAGCTAATCCAAATTAAGACACTGATGAACAAACACTCATGG 1159
QY 2846 aaagattactcgcagcgtctcagggagagctccatctgcagcggtctcgatcgcat 2905
Db 1160 CAGGATTACTACATATCTCTGGAGAACTTCATCTGCTGGAATTAATATGATCAT 1219
QY 2906 ttactgatatgtggcaaatlttgctgttccatgacacagcagttatggtggagtg 2965
Db 1220 TTCAGGTGATTTGGAGAC----- 1237
QY 2966 gacaaagaaatccaaattggcctctcgtgcagctagtgatgagaagaatalc 3025
Db 1238 -----TACATGCC 1246
QY 3026 caactatgatactgaatgtgctgtgacggtgggagaaacactataatgtggt 3085
Db 1247 AACTCGATGATCCGAGTACTGCTGCTGACTGATGGGAGATTAACATGCAAGTCTT 1306
QY 3086 gcttcaacgaggtcaacaaagtgtgccaatccacacagctgcttggggccctgt 3145
Db 1307 GTATTGATGAAGTGAAACAAAGTGGGCCATTGTTCAATTTTATTTGCTTGGGAAGACTG 1366
QY 3146 cagctcaagaactagagagcgtccaaatgacaggaagttacagacatgctcag 3205
Db 1367 CTGATGAACAGTAATTAAGATGATGAGCAAGTAACAGGAGAACTATTTTATGTTCTGAG 1426
QY 3206 atcaagttcagaacaaatggtcattgattgcttggggcccttcaacaggaatgtgag 3265
Db 1427 ATGAAGCTCAGAAATAGGCTCTATTGATCTTTGGGCTCTTACATVCAGAAATACCTG 1486
QY 3266 ctgtctctcagcgtccatccagcttgagagtaaggatlaaccctccagaacagccagt 3325
Db 1487 ATCTCTCCAGAGAGTCCCTTCACTGCAAGTAAGGATTAACACATGATATGATGCTT 1546
QY 3326 ggaatgaatggcagatgattgtgacagcagctgggaaaggaacttgttctatca 3385
Db 1547 GGATGAAACGACTGATTAATGATGATGATGAGGGAAGGAGCAACGCTTCTCTCATCA 1606
QY 3386 cctggagcaatgacgctccccaatctctctcgtggatcccaatgagcaagaagtg 3445
Db 1607 CATGGAACACTGCTCCCTCCAGTATTTCTGTGGGATCCCACTGGAACATTAATGAAA 1666
QY 3446 gctttagtggacaacaaatgagcctcactccaatcccaagattgtaag 3505
Db 1667 ATTTCACAGTGAATGCAACTTCCAAATGCGCTATCTCATATTTCCAGGAACATGCAAGG 1726
QY 3506 ttggcaacttggaaatacagctgc-----aagcaagctcacaacaccttgcacctgactg 3559
Db 1727 TGGGCACTTGGCATACATCTTCAAGCCCAAGCAACCCAGAAACATTAATATTATACAG 1786
QY 3560 tcaagctccgtgtgctcaaatgtaacctgacctcaatlaacagtgacttccaaaagaca 3619

Db 1787 TAATCTCTGAGCAGCAAAATTTCTGTGCTCCCATCATCAGATGCTAATAATGAATA 1846
QY 3620 agaacacagcaaatctccaccgctctggtatttataatgcaatattcgcaagagcct 3679
Db 1847 AGGACGTAAACAGTTCCTCCAGCCCATGATTTGTTTACGAGAAATTTCTCAAGGATATG 1906
QY 3680 ccccaatctcagggccagtgltcagcagccctgattgaaatcagtgaaatgaaacagtta 3739
Db 1907 TACCTGTTCTTGGAGCCCAATGCTACTCTTCAATGAAATCAGAAATGACATACAGAG 1966
QY 3740 ccttggacacacgttgaatgagcaggtgtgattgtatgtaataagagtgagtgact 3799
Db 1967 TTTTGGAACTTTTGGATAATGTGACAGGCGCTATTTCTTCAAGAAATGATGAGTACT 2026
QY 3800 caagatattccacaactatgacagatgtagatacaagtgaanaagtgcgggtctg 3859
Db 2027 CCAAGTATTTTACAGCATATACAGAAATGCGAGATATFACCTTAAAGTTGCGGCTCATG 2086
QY 3860 gaagagttaaagcagcagcagcagagtgatatacccaagagtgagacatgtacac 3919
Db 2087 GAGGAGCAAAACACTGCCAGGCTAAATTAACGGCTCCACTGAAATGAGCGCGCTACATAC 2146
QY 3920 ctgctggtatggaatgatatgaaatatacaatgaaatcccaagacctgaaatlaag 3979
Db 2147 CAGGCTGGGTAGTGAACGGGAAATTTGAACCAACCGCCAAAGCTGAATTTGAT--G 2203
QY 3980 atgattgtcacacaacagcagtggtgttccagcagaacatcctcgaggagctcattgtg 4039
Db 2204 AGGATACTGAGACCACTTGGAGGATTTTACGCCGAACAGCATCGGAGTGTGCTATTTGTG 2263
QY 4040 ctctcgtatgcccaaatgctccatcactgactcctccacactggtgcaaaatcacagacc 4099
Db 2264 TATCACAAGCTCCCAAGCTTCCCTTCCCTGACCAATACCAACCAAGCAATACAGACC 2323
QY 4100 tgaagcggaatctcaaggggagcagctcactaatctgacttggagagcgtcggggagt 4159
Db 2324 TTGATGCCACAGTTTCATGAGG---ATTAAGTATTTCTTACATGACACGCCAGGATTA 2380
QY 4160 attatgaccaatgaaacagctccacaagatataatcgaataaagtaacagatcttgatc 4219
Db 2381 ATTTTGTGTTGGAAAAGTTCAACGTTATATCATTAAGAAATAGTCGAAATGATCTTGATC 2440
QY 4220 tcaagagcaatgtaaatgtaactctctcaagtgaaatgtaactgctcactccaaaggaag 4279
Db 2441 TAAAGACACTTTTGAATGATGCTCTTCAAGTAATACTGATGATCTGACCAAGAGAGG 2500
QY 4280 ccaacttggagaagcttlttltttaaaccagaacaaacttacttggaaatggagcag 4339
Db 2501 CCAACTCCAAAGAAAGCTTTCATTTAAACAGAAATATCTCAGAAAGAAATGCAACC 2560
QY 4340 atcttctcatgtcatcagcgtgtgataagtcgactcgaatcagaataatccaca 4399
Db 2561 ACATATTATTTGTCATTAATAAGTATAGATTAAGCAATTTGACATTAATAAGTATCCACA 2620
QY 4400 ttgcagagatatttgttataatctccacagactccgcagaga---caactagtctg 4456
Db 2621 TTTCACAAGTAATCTTTGTTATTCCTCAAGCAAAATCCTGATGACATGATGACTACACTA 2680
QY 4457 atgaagcgtcgtccttgccttgcctaata 4483
Db 2681 CTCTACTCTCTACTCTCTACTCTGATA 2707

RESULT 9
AX092338
LOCUS AX092338 3265 bp DNA PAT 21-MAR-2001
DEFINITION Sequence 69 from Patent W00116318.
ACCESSION AX092338
VERSION AX092338.1 GI:13444483
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 3265)
AUTHORS Edton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
Wood,W.I.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0116318-A 69 08-MAR-2001;
Genentech, Inc. (US)
FEATURES
source location/Qualifiers
1..3265
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1159 a 596 c 632 g 878 t
ORIGIN
Query Match 24.7%; Score 1129.8; DB 6; Length 3265;
Best Local Similarity 67.1%; Pred. No. 1.1e-279;
Matches 1749; Conservative 0; Mismatches 747; Indels 111; Gaps 6;
QY 1889 aggcagatggtaccagcagcatctgtatctgtcttgaaagctaacaggaaagcatttatt 1948
DB 182 AGGATATGGTACACACCTTCTACGTACCTGTTGAAGCCAGAAAAGATTTTT 241
QY 1949 tcaaaaatgttgcattcttgatctcctgaacatgaaagcagctgactatgtgagac 2008
DB 242 TCAGAAATGTCTATATTAATTCCTGGAATTGGAGGAAATCCTCAGTACAAAGC 301
QY 2009 caaaactgagacctacaanaatgtctgattgtctgtgtctgagctactcctccagta 2068
DB 302 CAAGATGAAACCATTAACATGCTGATGTATGTGACACCACTCAGCTCCAGGTA 361
QY 2069 atgtgaaccttaacatgagatgagatgagcaactgtgagagaaggtgtaaaagatccacc 2128
DB 362 GAGATGAACCTACACCAAGCAAGTTCACAGATGTGAGAGAAAGCGAATATACATCTACT 421
QY 2129 tcaatcctgattcatctgcaagaaaagaagctgagctgaatgagcaagtaagagcat 2188
DB 422 TCACCCCGACCTTCTACTTGGAAAAAACAAATGATATGAGACCACAGCAAACTGT 481
QY 2189 ttgtccatgagtggtgctacatcagatgaggaagtaattgacgagtacaataatgata 2248
DB 482 TTGTCTCATGAGTGGGCTCAGCTCCGGTGGGAGGTGTGATGACGTAATGAGATCAGC 541
QY 2249 aatctactatccaatgagaagatac---aagcagtaagatgttcaacaggtattactg 2305
DB 542 CTTCCTACCGTCTAAGTCAAAAAAATCGAAGCAAGAGGTGTCGAGGTATCTCTG 601
QY 2306 gtacaatgtagtaagaagatgtcagggagcagctgttacaaccaaaagatgacattca 2365
DB 602 GTAGAAATAGAGTTTATTAAGTGTCAAGAGGAGCTGTCTTAGTGAACATGCAAAATTG 661
QY 2366 ataaagtaacagagactataigaaaaagaatgtagtltgttctccaatcccgccagagc 2425
DB 662 ATTCTACAGCAAACTGTATGGAAGATGTGCATTTCTTCTGTAAGTCAAAACAG 721
QY 2426 agaaggtctctataatgtttgacacacatgttgcattctatagttgaattctgtataagac 2485
DB 722 AAAAAGCTTCATATATGTTATGCAAGATTGATTGTTGTTGAATTTGTAAAGAA 781
QY 2486 aaaaaccacaagaagactccaacaagcaaatcaaaaatgcaatctccgaagacat 2545
DB 782 AAACCCATATCAGAAAGCTCCAGACCTCAACAAACATTAAGTGCATTTTGAATATACAT 841
QY 2546 gggaagtgatccgtgattctgaggaacttaagaanaacactctatgacaacacagcac 2605
DB 842 GGGAGGTATTAAGCAATCTGTGAGATTTTAAACACCATCCATCCATGGTGCACCACTTC 901
QY 2606 caaatcccaacttctcatgtctgagatgtgacaagaagatgtgtttagtccacttgaca 2665
DB 902 CTCACACCTGTCTTCATATGCTGGAAGATCAGTCAAGAAATGTGTGCTTAGTCTTTGATA 961

QY 2666 aatctgagaacatggtcgagctgtagtaccgcctccaatctcagctactcaacagcagcagctt 2725
DB 962 AGTCTGGAAGCATGGGGGCTTAAGACCGCCCTTAATTCGAATATCAACACACCAAAACATT 1021
QY 2726 tctgtctgcagacagcttgagcttggtgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2785
DB 1022 TCTGTCTGCACAGCTGTTGAATGATGATCTGTGGGATGTTCACTTGTATGTACTG 1081
QY 2786 cccatgtcaaaaatgaaactctacaagataaagaagtggaagtggaaggaacacactgcga 2845
DB 1082 CCACTATTGTAAATTAAGCTTAATCCAAATTAAGAGATGATGAAGAAACACACTCATAG 1141
QY 2846 aagaattacccgcagcagcttccaggaaggaagctccatctgcagcgggtctgcatcgcat 2905
DB 1142 CAGATTTACCTACATATTCCTCTGGGAGGAACTTCATCTGCTCGCATTAATATGAT 1201
QY 2906 ttaactgatatgtgcaaacatcttgcctgtttccatgacacacagcagatlatatgtgagatgc 2965
DB 1202 TTCAAGTGTATGAGAGC----- 1219
QY 2966 gacaagaataatccaatgtgggcctctcgccctgagcttagttagtaagaataatc 3025
DB 1220 -----TCAATTGCC 1228
QY 3026 caactgatatgctgaatgtgtctgtcagcagatggggaagacacataaagtgtg 3085
DB 1229 AACTCGATGAGATCCGAATACTGCTGTGACGTAGGGGAGATTAACCTGCAATTTCTT 1288
QY 3086 gcttgaacgagtcacaanaagttgtgtccatccacacagctgcgttggggccctctg 3145
DB 1289 GTATGATGAAGTGAACAAAGTGGGGCCATTTGATTTATTTCTTTGGGAAGAGCTG 1348
QY 3146 cagctcaagaactgagagagctgtccaaaatgacaggggtttaagacataatgtctag 3205
DB 1349 CTGATGAAGCAATTAATTAAGATGAGACCAAGATTAACGGAGAGTCAATTTATGTTTAG 1408
QY 3206 atcaagttcagaacaatgtgctcattgtatgtcttggggcccttcacacagaaatgag 3265
DB 1409 ATGAAGCTCAGAACAAATGGCCCATTTGATGCTTTGGGGCTTTACATCAGGAATTAAG 1468
QY 3266 ctgtctctcagcgtccatccagcttgagagtaaggatlaacccctccagaacagcagct 3325
DB 1469 ATCTCTCCAGAGCTCCCTTCAGCTCGAAGTAAGGATTAACATGTAATGATGCTT 1528
QY 3326 ggaatgaagacagatgtagtgcgagacacgcgtggaagaagacattgtttctatca 3385
DB 1529 GGATGAACGACCTGTCTAATTAATGATGATGACGTGGAAAGGACAGCTTCTTCATCA 1588
QY 3386 cctggaacaatgacgctccccaatcccttctctgtggaatcccgatggaacagaagctg 3445
DB 1589 CATGGAACAGTCTGCTCCAGATATTCTCTCTGTGGATCCCGATGGAACAAATTAAGTGA 1648
QY 3446 gctttagtgtagacaanaacaccaaaatggtcctacatcccaatcccgagcatctgaag 3505
DB 1649 ATTTCAGATGATCAACTTCCAAATAGGCTTATCTAGTATTCAGAGAACATGCAAGG 1708
QY 3506 ttggcattggaataatcagctgcg-----aagcagcttcaaaaacttgcacccgagct 3559
DB 1709 TGGGCATTGGGCATACATCTTCAAGCCAAAGCCCAACCCCAACAACTAATTAATTACG 1768
QY 3560 tcaagctcccgtagtcccaatgctacccgtcccaatlaagatgacttccaagaagaaca 3619
DB 1769 TAACTTTCGAGAACCAATTTCTGTGCGCTCCCAATCAGAGTGAATGTAATTAATGAATA 1828
QY 3620 aggaacacagaacaaatccccagcctctgtgattatgcaaatatctgcgaagagcct 3679
DB 1829 AGGAGCTAAACAGTTTCCCGCCCAATGATGTTTACGAGAAATTTCTACAAAGATATG 1888
QY 3680 ccccaactctgaagcagatgtcagagcccgatggaatgaatgaatgaatgaatgaatga 3739
DB 1889 TACCTGTTCTTGGAGCCATATGACTGCTTTCATTTGAATTCAGAGATGACATTAAGAG 1948
QY 3740 ccttgaactactgataatggaacaggtgtctgattactactaagaatgaacggtgtctact 3799

```

1949 TTTTGGAACTTTTGGATTAAGTGTCAGGCGCTGATTTCTTTCAAGAAATGAGTCTACT 2008
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3800 caaggtttttcacaaactatgacacgagtgatagatcacgttgaaagtcgggctctg 3859
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2009 CCAAGTATTATTACAGCATATACAGAAATGCGACATATACCTTAAAGTTGCGGCTATG 2068
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3860 gaagagttacacgacgacgacgagagtgatagatccacgacgagtgagacactgtacat 3919
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2069 GAGGAGCAACAGCTGCGAGGCTAAATATTACGGCTCCACCTGAATATAGACCGCGGTACATAC 2128
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3920 ctggcttgatgtgagatgagatgaaatacaatgtaacccaagaacctgaaataaagg 3979
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2129 CAGGCTGGAGTGAACGGGGAAATTGAACCAACCCGCAAGCCCTGAATATTGAT---G 2185
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3980 atgatgtcacacaaacgacgagtggtgttttcagcagaacatctcgggagggcatttgg 4039
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2186 AGGATCTACAGACCACTTGGAGAGATTTCAGCCGACACAGCATCCGAGGTGCTATTGTGG 2245
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4040 ctctcgtatgcccaaatgctccatatactctctcccaactgacaaatcacgacc 4099
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2246 TATCACAAGTCCCAAGCCTTCCCTTGCTGACCAATATCCACCAAGTCAATATCAGAGCC 2305
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4100 tgaagtgagaaattcaacgggggagctctcattatctgacttggacagctccctgggagt 4159
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2306 TTGATGCCACAGTTCATGAGG---ATTAAGATTATTCTTACATGAGACACACAGAGATA 2362
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4160 attatgacccatgaaacgacgacacgaatatactcgaataagtaacagatcttgatc 4219
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2363 ATTTTGATGTTGAAAGTTCAACGTTATATATCATTAAGATAATAGTCAAGTATTTCTTATC 2422
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4220 tcagagacaaatgtaaatctctcctcaagtgaatactactgctctcatcccaaggag 4279
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2423 TAAGAGACAGTTTGAATGATGATCTTCAAGTAATACTGATGATCTGTCACCAAGAGAGG 2482
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4280 ccaactctgagagagctcttctttaaaccagaaacattactcttgaagaatggcag 4339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2483 CCAACTCCAGGAAGGCTTTTGCAATTTAAACCAAGAAATATCTCAGAAAGAAATGCAACCC 2542
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4340 atcttctcatgctcattcagcgctgtgataagtgatcgtgaatcagaataatcacaca 4399
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2543 ACATATTATTGTCATTAAAGTATAGATTAAGCAATTGGACATCAAAAGTATCCACACA 2602
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4400 tgcacgagatcttgatttattctccacagactccgcacagaga---cacctagtcctg 4456
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2603 TTGCACACAGTAACCTTTGTTATCCCTCAACCAATCTGATGACATTGATCCTACACCTA 2662
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4457 atgaagctctgctctctgtcctaata 4483
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2663 CTCCTACTCTACTCTACTCTCTGATA 2689
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 10
LOCUS AC068071/c 175591 bp DNA HTG 28-MAY-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-290N3, WORKING DRAFT SEQUENCE,
15 unordered pieces.
ACCESSION AC068071
VERSION AC068071.4 GI:8099747
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE 1 (bases 1 to 175591)
AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Federespiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175591)
AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Federespiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,

```

TITLE

JOURNAL

COMMENT

Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
 Submitted (28-APR-2000) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 On May 28, 2000 this sequence version replaced gi:7671256.
 ----- Genome Center
 Center: Stanford DNA Sequencing and Technology Development
 Center

Center code: SDSRDC

Web site: <http://sequence-www.stanford.edu/group/human/>
 Contact: hum-info@sequence.stanford.edu

Project Information
 Center project name: 889
 Center clone name: RP11-290N3

----- Summary Statistics

Sequencing Vector: M13mp18; X02513
 Chemistry: Dye-Primer; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319
 Consensus quality: 160064 bases at least Q40

Consensus quality: 168051 bases at least Q30
 Consensus quality: 170998 bases at least Q20

Insert size: 161152; agarose-fp
 Insert size: 178431; sum-of-contigs

Quality coverage: 4.3x in Q20 bases; agarose-fp
 Quality coverage: 3.9x in Q20 bases; sum-of-contigs.

* NOTE: This is a 'working draft' sequence. It currently
 consists of 15 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.

* as soon as it is available and the accession number will
 be preserved.

```

1 1013: contig of 1013 bp in length
* 1014 1113: gap of unknown length
* 1114 2493: contig of 1380 bp in length
* 2494 2593: gap of unknown length
* 2594 4311: contig of 1718 bp in length
* 4312 4411: gap of unknown length
* 4412 8684: contig of 4273 bp in length
* 8685 8784: gap of unknown length
* 8785 13051: contig of 4267 bp in length
* 13052 13151: gap of unknown length
* 13152 22390: contig of 9239 bp in length
* 22391 22490: gap of unknown length
* 22491 34595: contig of 12105 bp in length
* 34596 34695: gap of unknown length
* 34696 46780: contig of 12085 bp in length
* 46781 46880: gap of unknown length
* 46881 59460: contig of 12580 bp in length
* 59461 59560: gap of unknown length
* 59561 74872: contig of 15312 bp in length
* 74873 74972: gap of unknown length
* 74973 89920: contig of 14948 bp in length
* 89921 90020: gap of unknown length
* 90021 107986: contig of 17976 bp in length
* 107987 108096: gap of unknown length
* 108097 129557: contig of 21461 bp in length
* 129558 129657: gap of unknown length
* 129658 150564: contig of 20907 bp in length
* 150565 150664: gap of unknown length
* 150665 175591: contig of 24927 bp in length.

```

FEATURES

source

1. 175591

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/clone="RP11-290N3"

/clone_lib="RPC1 human BAC library 11"

misc_feature

1. 1013

```

misc_feature /note="assembly_name:Contig8"
1114..2493
misc_feature /note="assembly_name:Contig10"
2594..4311
misc_feature /note="assembly_name:Contig11"
4412..8684
misc_feature /note="assembly_name:Contig12"
8785..13051
misc_feature /note="assembly_name:Contig13"
13152..22390
misc_feature /note="assembly_name:Contig14"
clone_end:SP6
vector_side:left"
22491..34595
misc_feature /note="assembly_name:Contig15"
34696..46780
misc_feature /note="assembly_name:Contig16"
46881..59460
misc_feature /note="assembly_name:Contig17"
59561..74872
misc_feature /note="assembly_name:Contig18"
74973..89920
misc_feature /note="assembly_name:Contig19"
90021..107996
misc_feature /note="assembly_name:Contig20"
108097..129557
misc_feature /note="assembly_name:Contig21"
129658..150564
misc_feature /note="assembly_name:Contig22"
150665..175591
misc_feature /note="assembly_name:Contig23"
BASE COUNT 53213 a 32926 c 33415 g 54576 t 1461 others
ORIGIN

```

```

Query Match 23.2%; Score 1058.2; DB 2; Length 175591;
Best Local Similarity 95.3%; Pred. No. 4.6e-264;
Matches 1125; Conservative 0; Mismatches 2; Indels 54; Gaps 1;

QY 1 atgataaatctcctacacatacactctcccaagactaaccaggaagtgaatct 60
DB 62439 ATGGATTAATTCCTCAACACATACACTCTCCCAAGACTAAACAGAGAGTAATCT 62380

QY 61 ctgatatagccaataaaggagctctgatatcttggaataataaaggctttaccaccaa 120
DB 62379 CTGAAATAGACCAATATACAGGCTCTGATATTGTGGCAATATACAGAGCTTACCAACCA 62320

QY 121 aagagtcaggagaccagatgattccacagctgaattctaccagaggtacaaggaggaactg 180
DB 62319 AAGAGTCCAGAGACCGAGATGATTCACAGCTGAATCTACAGAGGTACAGAGAGAACTG 62260

QY 181 gtaccatctcctctgaaagtattacaatcagaagaaagggaactcctcctaactcg 240
DB 62259 GTACCATTCCCTCTGAAAGTATTACAAATCAATGAAAAAGGCAATCCTCCTTAACCTG 62200

QY 241 ttatatgggccaacatctctgtatccaaaggcgggcagagacacaaacaaaggag 300
DB 62199 TTTTATGGGCGCAACATATCTGTATACCAAAAGCGGGCAGAGACACAAACAAAAAGAG 62140

QY 301 aatttagaccaatacttgatgacacatgtagcaaaaatcctcctaataaatactggca 360
DB 62139 AATTTTAGACCAATATCTTTGATGAACATGTGATGCAAAAATCCTCAATTAATTAATGCGA 62080

QY 361 aacgaatccagcagacacataaaaggcttatccacatgatcaagtggtcctccct 420
DB 62079 AACGAATCCAGCAGACACATCAAAAAGCTTATCCACATGATCAAGTGGGCTTCACTCCT 62020

QY 421 gggat-----a 426
DB 62019 GGGATGTAAGGCTGTCAATATACGCAATCAATAAATGTAATCCAGCATATAACAGCA 61960

QY 427 accaaagcaaaaacacatgatattctcaatagatgcagaaaaggccttggacaaatt 486

```

```

DB 61959 ACCAAAGACAAAACCATGATATCTCAATAGATGAGAAAAGCCTTTGACAAAATT 61900
QY 487 caacacaccttatgtctataaaacctcaataaattagatatgtatgggacatatctcaa 546
DB 61899 CAACACACCTTCATCTATAAAACCCCTCAATAAATTGATTTGAGGGACATATCTCAA 61840
QY 547 ataataagagctatctatggaagccaagccaagcacaatcatatcatatgagggaaaaactg 606
DB 61839 ATAATAGAGCTATCTATGCGCAAAACCCAGCCATATCATATCATAGGGGCAAAAACCTG 61780
QY 607 gaagcattcccttggaaaacttggcacaagacagagagtgcctctctccaccatctatlc 666
DB 61779 GAAGCATTCCTCTTGAAAACCTGGCACAAGAGGATGCGCTCTGCAACCATCTCTATTC 61720
QY 667 aacatagttttggaaagttctggccaaggaacttaaggcagagaagaaataagggtttt 726
DB 61719 AACATAGTTTGGAAAGTTCTGCGCAGGGCAATTAAGCCACAGCAATAAATAAAGGCTTTT 61660
QY 727 caattaggaagaaagagatcaaatgttccctgtttgcaagtgatgatgtatlatccta 786
DB 61659 CAATTAGCAAAAAGAGAAAGTCAAAATTGTCCTGTTTGCAGGTGACATGATTTGTAACCTA 61600
QY 787 gaaaaccccatctctcagcccaaaatctccttaagctgataagaaacttcagcaaaagtc 846
DB 61599 GAAAACCCCATCTCTCAGCCCAAAAATCTCTTAAGCTGATAGCAACTTCAGCAAAAGTC 61540
QY 847 tcaggatccaaaatcaatgltacaaaataacagacattcctctatcaccaataacagagaa 906
DB 61539 TCAGATATCAAAATCAATAGTACAAAATCACAAGATTCCTTATACCAATATACAGAGAA 61480
QY 907 acagagagccaatatatgaatgaactcccatcacatgtctcacaagagaataaataac 966
DB 61479 ACAGAGAGCCAAATATATGATATATCACTCCCATTCACAAATTTCTTCAAGAGATTAATATC 61420
QY 967 ctaggaaatccaaacttacaagagatgtgaagacactcttcaaggagaaactacaacacactg 1026
DB 61419 CTAGGAATCCAACTTACAAAGGATGTGAAGACCTCTTCAAGAGAACTACAAACACACTG 61360
QY 1027 ctcaatgaataaagaagatacaaaaataaalgaaagaacattcgcatactgataagaa 1086
DB 61359 CTCATATGAATTAAGAGAGATCAAAACAAATGGAAGAACATTCATGCTCATGATGATGCA 61300
QY 1087 agaatacatatcgtgaaaatggccatctgacacagatatt 1127
DB 61299 AGAATCATATGCTGAATAATGGCCATTAAGTCCCAAGGTAAT 61259

```

```

RESULT 11
AL356270 140718 bp DNA HTG 23-JAN-2001
LOCUS Homo sapiens chromosome 1 clone RP5-873015, *** SEQUENCING IN
DEFINITION
PROGRESS ***, 22 unordered pieces.
ACCESSION AL356270.3 GI:9797470
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 140718)
AUTHORS Pavitt,R.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:9213800.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: d1873015

```

```
----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 12% of reads
Sequencing vector: plasmid; L08752; 87% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 129686 bases at least Q40
Consensus quality: 134097 bases at least Q30
Consensus quality: 136445 bases at least Q20
Insert size: 138618; sum-of-contigs
Quality coverage: 3.10x in Q20 bases; sum-of-contigs quality
coverage: 3.21x in Q20 bases; agarose-1f

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3292: contig of 3292 bp in length
* 3293 3392: gap of 100 bp
* 3393 10168: contig of 6776 bp in length
* 10169 10268: gap of 100 bp
* 10269 21700: contig of 11432 bp in length
* 21701 21800: gap of 100 bp
* 21801 29662: contig of 7862 bp in length
* 29663 29762: gap of 100 bp
* 29763 31862: contig of 2100 bp in length
* 31863 31962: gap of 100 bp
* 31963 36577: contig of 4615 bp in length
* 36578 36677: gap of 100 bp
* 36678 43943: contig of 7266 bp in length
* 43944 44043: gap of 100 bp
* 44044 52906: contig of 8863 bp in length
* 52907 53006: gap of 100 bp
* 53007 65117: contig of 12111 bp in length
* 65118 65217: gap of 100 bp
* 65218 70699: contig of 5482 bp in length
* 70700 70799: gap of 100 bp
* 70800 82141: contig of 11342 bp in length
* 82142 82241: gap of 100 bp
* 82242 87387: contig of 5146 bp in length
* 87388 87487: gap of 100 bp
* 87488 91496: contig of 4009 bp in length
* 91497 91596: gap of 100 bp
* 91597 104147: contig of 12551 bp in length
* 104148 104247: gap of 100 bp
* 104248 107008: contig of 2761 bp in length
* 107009 107108: gap of 100 bp
* 107109 110463: contig of 3355 bp in length
* 110464 110563: gap of 100 bp
* 110564 112670: contig of 2107 bp in length
* 112671 112770: gap of 100 bp
* 112771 116165: contig of 3395 bp in length
* 116166 116265: gap of 100 bp
* 116266 118389: contig of 2124 bp in length
* 118390 118489: gap of 100 bp
* 118490 129046: contig of 10557 bp in length
* 129047 129146: gap of 100 bp
* 129147 136534: contig of 7388 bp in length
* 136535 136634: gap of 100 bp
* 136635 140718: contig of 4084 bp in length.

FEATURES
source
1. 140718
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="1"
  /clone="RP5-873015"
  /clone_lib="RPCT-5"
  /note="assembly_fragment:00102
misc_feature
1. 3292
/note="assembly_fragment:00102
```

```
fragment_chain:1"
3393. 10168
/note="assembly_fragment:00109
fragment_chain:1"
10269. 21700
/note="assembly_fragment:00228
fragment_chain:2"
21801. 29662
/note="assembly_fragment:00045
fragment_chain:2"
29763. 31862
/note="assembly_fragment:00277
fragment_chain:3"
31963. 36577
/note="assembly_fragment:00849
fragment_chain:3"
36678. 43943
/note="assembly_fragment:00578
fragment_chain:4"
44044. 52906
/note="assembly_fragment:00490
fragment_chain:4"
53007. 65117
/note="assembly_fragment:01028
fragment_chain:5"
65218. 70699
/note="assembly_fragment:00015
fragment_chain:5"
70800. 82141
/note="assembly_fragment:00071"
82242. 87387
/note="assembly_fragment:00242"
87488. 91496
/note="assembly_fragment:00323"
91597. 104147
/note="assembly_fragment:00430"
104248. 107008
/note="assembly_fragment:00435"
107109. 110463
/note="assembly_fragment:00756"
110564. 112670
/note="assembly_fragment:00793"
112771. 116165
/note="assembly_fragment:00810"
116266. 118389
/note="assembly_fragment:00912"
118490. 129046
/note="assembly_fragment:01001"
129147. 136534
/note="assembly_fragment:01002"
136635. 140718
/note="assembly_fragment:01014"

BASE COUNT 44522 a 26618 c 25494 g 41976 t 2108 others
ORIGIN
Query Match 23.0%; Score 1051.8; DB 2; Length 140718;
Best Local Similarity 94.8%; Pred. No. 2e-25;
Matches 1120; Conservative 0; Mismatches 7; Indels 54; Gaps 1;

QY 1 atgataaattcctcaacatacaactctccaagactaaacaggaagaagtgaattc 60
|||||
Db 105397 ATGGAATTAATTCCTCGACACATFACACCTCCCAAGACTAAACAGAGAGTTGAATCT 105456.

QY 61 ctgaatagaccataacagagctctgatattgtygcaataatcaagagcttaccaccaa 120
|||||
Db 105457 CTGAATAGACCAATAACAGGCTCTGATATTGTGCAATATCAAGAGCTTACCAACCAA 105516

QY 121 aagagtcacagacagatgatattcaagctgaattctaccacagaggttaccagaggaact 180
|||||
Db 105517 AAGAGTCCAGACACAGATGAGATTACAGCTGAATTCTACCAAGAGGTACAGAGAACTG 105576

QY 181 gtaccattccctctgaagaatatacaatacaagaagaaggaactctccctaactcg 240
```



```
repeat_region 13912..14264
/note="LI repeat: matches 3591..3955 of consensus"
repeat_region 14260..14575
/note="LI repeat: matches 4002..4324 of consensus"
repeat_region 14582..15269
/note="LIM47 repeat: matches 596..1 of consensus"
repeat_region 15127..15415
/note="LI repeat: matches 5390..5093 of consensus"
prim_transcript <151414..>155920
/note="match: 5' EST R19063 clone 33638; match: 5' EST
H21576 clone 160009"
repeat_region 16063..16270
/note="LI repeat: matches 5083..4876 of consensus"
repeat_region 16327..16746
/note="MST4 repeat: matches 1..426 of consensus"
repeat_region 16747..17156
/note="MST-INTERNAL repeat: matches 1..413 of consensus"
repeat_region 17152..17493
/note="LTR7 repeat: matches 450..100 of consensus"
repeat_region 17410..17534
/note="LTR7 repeat: matches 131..1 of consensus"
prim_transcript <20771..21634
/note="match: multiple ESTs; match: 3' EST H60808 clone
208209; match: 5' EST H666214 clone 234021; paired with
EST H666215 matching this clone; match: 3' EST H666215
clone 234021; paired with EST H666214 matching this clone;
match: 5' EST H59570 clone 206711; paired with EST H59571
matching this clone; match: 3' EST H59571 clone 206711;
paired with EST H59570 matching this clone; match: 3' EST
H66951 clone 295823; match: EST T12379 clone A151"
prim_transcript complement(22040..22488)
/note="match: EST H60808 clone 28e05; similar to RTVL-H
related sequence"
repeat_region 22355..22757
/note="LTR7 repeat: matches 450..1 of consensus"
repeat_region 22766..23422
/note="MST-INTERNAL repeat: matches 408..1075 of
consensus"
repeat_region 23525..23740
/note="MSTD repeat: matches 1..216 of consensus"
repeat_region 23739..23847
/note="MSTC repeat: matches 294..402 of consensus"
repeat_region 24176..24211
/note="18 copies of 2 mer 89 & conserved"
repeat_region 24210..24284
/note="3 copies of 25 mer 83 & conserved"
repeat_region 24262..24297
/note="18 copies of 2 mer 83 & conserved"
repeat_region 24300..24337
/note="19 copies of 2 mer 84 & conserved"
repeat_region 24826..24853
/note="14 copies of 2 mer 96 & conserved"
repeat_region 24858..25456
/note="LI repeat: matches 4213..3608 of consensus"
repeat_region 25471..26788
/note="LI repeat: matches 3487..2133 of consensus"
repeat_region 26926..26977
/note="26 copies of 2 mer 83 & conserved"
repeat_region 28145..28279
/note="FLAM C repeat: matches 5..133 of consensus"
repeat_region 28711..28760
/note="LIM42 repeat: matches 1005..1055 of consensus"
repeat_region 28813..28998
/note="LIP15 repeat: matches 904..718 of consensus"
repeat_region 28996..29888
/note="LIP2 repeat: matches 893..1 of consensus"
repeat_region 29739..34240
/note="LI repeat: matches 5390..896 of consensus"
repeat_region 34236..35009
/note="LI repeat: matches 772..8 of consensus"
repeat_region 35015..35737
/note="LIP9 repeat: matches 741..1 of consensus"
repeat_region 35606..36916
```

```
/note="LI repeat: matches 5390..4057 of consensus"
36911..38329
/note="LI repeat: matches 3761..2341 of consensus"
repeat_region 38335..38586
/note="LI repeat: matches 2230..1966 of consensus"
repeat_region 38587..38892
/note="LIS9 repeat: matches 303..1 of consensus"
repeat_region 38893..39448
/note="LI repeat: matches 1977..1430 of consensus"
repeat_region 39243..39526
/note="MER25 repeat: matches 2136..1844 of consensus"
repeat_region 39618..40048
/note="LIP7 repeat: matches 886..464 of consensus"
repeat_region 42147..42518
/note="MLT1B repeat: matches 11..374 of consensus"
repeat_region 42532..42565
/note="17 copies of 2 mer 88 & conserved"
repeat_region 44113..44162
/note="25 copies of 2 mer 96 & conserved"
repeat_region 44955..47064
/note="LI repeat: matches 990..3058 of consensus"
repeat_region 47136..48331
/note="LI repeat: matches 4176..5390 of consensus"
repeat_region 48187..49016
/note="LIM4 repeat: matches 1..845 of consensus"
repeat_region 49038..49921
/note="LIP14 repeat: matches 895..1 of consensus"
repeat_region 49772..50034
/note="LI repeat: matches 5390..5129 of consensus"
repeat_region 50037..50318
/note="LI repeat: matches 4864..5141 of consensus"
repeat_region 50300..50507
/note="LIM45 repeat: matches 832..1045 of consensus"
repeat_region 50562..50780
/note="LIM3 repeat: matches 231..3 of consensus"
repeat_region 51075..51929
/note="LIM10 repeat: matches 1069..234 of consensus"
prim_transcript <52519..>53177
/note="match: multiple ESTs; match: 5' EST H87217 clone
220533; match: 5' EST R05992 clone 125327; match: 5' EST
R99371 clone 201201; match: 3' EST T64484 clone 80506;
match: 3' EST R42572 clone 30927; match: 3' EST R42487
clone 31053"
repeat_region 55349..55378
/note="15 copies of 2 mer 90 & conserved"
repeat_region 56474..56523
/note="25 copies of 2 mer 96 & conserved"
repeat_region 56550..56575
/note="13 copies of 2 mer 92 & conserved"
repeat_region 56780..57164
/note="MST4 repeat: matches 1..426 of consensus"

Query Match 23.0%; Score 1049.4; DB 9; Length 118426;
Best Local Similarity 80.9%; Pred. No. 8e-25;
Matches 1364; Conservative 0; Mismatches 166; Indels 155; Gaps 6;

QY 1 atgataaattcctcaacacatacactctccaaagctaaacaggaagtgtatct 60
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 105431 ACGGATTAATTCTCGGACATATACCGTCCCAAGCAAGCAAGCAAGCAATCC 105430

QY 61 ctgaatagaccaataacaggctctgatattgtgcataatacaagagcttaccaccaaa 120
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 105491 CTGAATAGACCAATTAACAGTCTGAATGAGCGCTATTAACGCCATCAACCAAA 105550

QY 121 aaagatccaggacagatgatcatcagctgaatttaccagggtacaaaggaggaacg 180
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 105551 AAAAGTCCAGGACCAAGCAATTCACAGCCATTCTACCAAGGATCAAGAGCTG 105610

QY 181 gtaccattccctcgaagtattacatacaatagaanaaggaatccctcctaactg 240
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 105611 GTACCA-TCCCTTGTGAACATTTCCAAACATATGAAGAAGAGGATCCCTCACTCA 105669

QY 241 ttattatgagccaacatcctgataccaagaacggcgagagacacacccaanaagag 300
```

```

|||||
Db 105670 TTTATGAGCCAGCATCATCTATACCCAAACCTGCGACAGACACAAACAAAAAGAA 105729
QY 301 aattttagccaatatctttgatgaacatltgatgaacaaatccctcaataaatactgtgca 360
Db 105730 AATTTCAGGTCAATTAATTCCTATGATCACTGATGCAAAAAATCCCAATAAATACTGGCA 105789
QY 361 aacgaatccagcagcacatcaaaaagcttaccaccaatgaatgaatgtggcttccct 420
Db 105790 AACCAATCCAGCAGCAGCATCAAAAAGCTTATACATCGATCAAGTTGGCTTCCTCCT 105849
QY 421 ggaat-----a 426
Db 105850 GGGATGCAAACTGGTTCACAAATGTAATTAATTAATTAATTCATCATACATAACAGA 105909
QY 427 accaaagccaaacccaatgatattatcctcaatagatgagaaagccttggcaaaatt 486
Db 105910 ACCAATTAACAAACCAACAAAGATTAATTCATAGTACAGAAAAGCCCTTCACAGCAAT 105969
QY 487 caacaacccctcatgctaaaaacccctcaataatagatatgtatggagacatctcaaa 546
Db 105970 CAACACCCCTTCATGCTAAAAAATCTCAATTAATTAATTAATTAATTAATTAATTAAT 106029
QY 547 ataataagagctatctatgcaaaagccacagccaatatcactgaatlyggcaaaactg 606
Db 106030 ATAATTAAGAGCTATTTATGACAAACCCACAGCAATATCATCTAGTAAGGCAAAAAC 106089
QY 607 gaagca-ttccctttgaaactgtgcaaaagcaggaatgacccctctcctcaactctatt 665
Db 106090 GAACGAGATCCCTTTGATGATGCGACAAA----GGATGCCCTCTCTACACATCCCTAT 106144
QY 666 caacatagttcttggaagttctgcccaggaagcattagcaggagaaggaataaaagttc 725
Db 106145 CAACATATATTTGGAAGTTCTGGCCAGGCAATCAGAAAGAGAGAGAAATTAAGGAT 106204
QY 726 tcaatlaggaagaaaggaagatcaaatlytccctgttgcagagtgacatgattatcct 785
Db 106205 TCAATATAGGAAGAGAGAGAAATCCCTCTCTGTTGCAATGACATGATTTGATATTT 106264
QY 786 agaaaaacccctctctctagcccaaaatctccttaagctgtatgaagcaactcagaagct 845
Db 106285 AGAAAAACCCCATCGTCGAGCCCAAAATCTCTTAAGCTGATTAACCAATTCAGCAAGT 106324
QY 846 ctcaagatacaaatcaatgtacaaaatcaaaagcatctcctatacacaataacagaga 905
Db 106325 CTCAGGATACAAATCATATGTCGCAAAATTCACAAAGCATTCCTATACCAATTAACAGCA 106384
QY 906 aacagagagccaaatcatgaatgaatcccatcacaattgcttcaaaagagaataaata 965
Db 106385 AACAGAGAGCCAAATCATGATGAATCCCATTCCTAATTTGCTCAAGAGAAATTAATA 106444
QY 966 cctaggaaatccaacttacaagagatgtgaaggaactcttcaagaggaactcaaacact 1025
Db 106445 CCTATGAAATTAACCTTAACG----GAAGGACCTCTTAAGGACAACTCAAAACCACT 106499
QY 1026 gctcaatgaataaaagagagatacaacaatgtgaagacattcactgtcatgatatgag 1085
Db 106500 GCACAGGAATAATTAAGAGAGACACAAACAAATGCAAAAAGATTTCAATGCTCATGATATG 106559
QY 1086 aagaatcaatctgtgaataatgagccatctgcccagaattatgtatagatataaaggtat 1145
Db 106560 AAGATCAATATTTGGAATAATGCGCATACTGCCCAAACTAATTTACCAATTAATGCTAT 106619
QY 1146 ----- 1145
Db 106620 CTCATCAAGCTACCACTGACTTTCTTCAAGAACTGCAAAAACCTACTTAAGCTCAT 106679
QY 1146 -----tcaatlaggaagaggaagtgcaaatgtc 1176
Db 106680 ATGGAACCAAAAAAGAGCGCGCATAGCAAGACAAATTTCTTAAGCAAAATGTCAAAATGTCT 106739
QY 1177 ctgttgcagatgacatgattgtatatctagaaaaacccatgtctcagcccaaatctc 1236
|||||

```

```

Db 106740 CTGTTTCAGATGACATGATTTGATATTTAGAAAAACCCAGGCTCTCAACTCAAAATCTT 106799
QY 1237 cttaagctgataagcaacttaagcaagctctcagagatacaaatcaatgtacaaaatca 1296
Db 106800 CTTAAGCTGATTAAGAACTTACGCAAAAGTCTCAGATACCAAAATCAATGTCAAAAATFCA 106859
QY 1297 caagcatctctacacacacaacagacaacagagagccaaatcaatgaatgtgaactccca 1356
Db 106860 CAAGCATTTCTATACCAATTAATTAACAAACAGAGACCAAAATCATGAGTAAGTCCCA 106919
QY 1357 ttcaaatgtcttcaaaagataaataatcactaggaatcccaacttacaagagcgtgaag 1416
Db 106920 TTCACAATTTGCTACAAAGATAAATGCTATCAATTAACAACTTACAAAGATGTGAAG 106979
QY 1417 gacctctcaagaggaactacaacacactgctcaaggaataaagaggtatacaaaaaa 1476
Db 106980 GACCTCTTCAAGAGAACTACAAACCACTGCTCAAGGAAATTAAGAGGAAACAAACAAA 107039
QY 1477 tggagaacattctcatgtcatgataagaaatcaatcactgtgaatgtgcaactatg 1536
Db 107040 TGGAATAACCATTCATGCTCATGATAGAGAAATCATATCTGTAGAGAGGCCATCTG 107099
QY 1537 cccaa 1541
Db 107100 CCCAA 107104

RESULT 13
AC010275/c
LOCUS
DEFINITION
Homo sapiens chromosome 5 clone CTC-499J9, WORKING DRAFT SEQUENCE,
15 ordered pieces.
AC010275
AC010275.5 GI:9256179
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEPIN.
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 196805)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 196805)
DOE Joint Genome Institute.
Direct Submission
Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
on Jul 18, 2000 this sequence version replaced gi:7710793.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 377021
Center clone name: CIT-HSPC_499J9
-----
Summary Statistics
Consensus quality: 185697 bases at least Q40
Consensus quality: 193135 bases at least Q30
Consensus quality: 194420 bases at least Q20
Estimated insert size: 194000; pulse field gel estimation
Estimated insert size: 196105; sum-of-contigs estimation
Quality coverage: 6.0 in Q20 bases; pulse field gel estimation
Quality coverage: 5.93 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced

```


NOTE: This is a "working draft" sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
misc_feature 1. .1215
              /note="assembly_name:Contig10
              clone_end:T7
              vector_side:left"
```

misc_feature 1316..2643 /note="assembly_name:Contig11"
misc_feature 2744..4582 /note="assembly_name:Contig12"
misc_feature 4683..6572 /note="assembly_name:Contig13"
misc_feature 6673..8059 /note="assembly_name:Contig14"
misc_feature 8160..9839 /note="assembly_name:Contig15"
misc_feature 9940..11957 /note="assembly_name:Contig16"
clone_end:SP6
vector_side:right"
misc_feature 12058..14675 /note="assembly_name:Contig17"
misc_feature 14776..16933 /note="assembly_name:Contig18"
misc_feature 17034..19985 /note="assembly_name:Contig19"
misc_feature 20086..24485 /note="assembly_name:Contig20"
misc_feature 24586..28286 /note="assembly_name:Contig21"
misc_feature 28387..32573 /note="assembly_name:Contig22"
misc_feature 32674..36708 /note="assembly_name:Contig23"
misc_feature 36809..41692 /note="assembly_name:Contig24"
misc_feature 41793..47698 /note="assembly_name:Contig25"
misc_feature 47799..54305 /note="assembly_name:Contig26"
misc_feature 54406..61109 /note="assembly_name:Contig27"
misc_feature 61210..68598 /note="assembly_name:Contig28"
misc_feature 68699..77839 /note="assembly_name:Contig29"
misc_feature 77940..85716 /note="assembly_name:Contig30"
misc_feature 85817..93966 /note="assembly_name:Contig31"
misc_feature 94067..103047 /note="assembly_name:Contig32"
misc_feature 103148..113126 /note="assembly_name:Contig33"
misc_feature 113227..122160 /note="assembly_name:Contig34"
misc_feature 122261..135587 /note="assembly_name:Contig35"
misc_feature 135688..147635 /note="assembly_name:Contig36"
misc_feature 147736..162548 /note="assembly_name:Contig37"
misc_feature 162649..163761 /note="assembly_name:Contig4"
misc_feature 163862..165252 /note="assembly_name:Contig7"
misc_feature 165353..166804 /note="assembly_name:Contig8"
misc_feature 166905..168185 /note="assembly_name:Contig9"
BASE COUNT 50525 a 30701 c 29699 g 54153 t 3107 others
ORIGIN

Query Match 22.4%; Score 1024.8; DB 2; Length 168185;
Best Local Similarity 80.8%; Pred. No. 1.8e-252;
Matches 1362; Conservative 0; Mismatches 167; Indels 157; Gaps 8;

QY 1 atgataaattcctcaacatacactctcccaagactaaaccagaagaagtgaattc 60

|||||
Db 124416 ACGGATAAATTCCTGGACACATATACCGTCCCAAGAGAAACAAAGGAAGTCAATATCC 124475
QY 61 ctgaatagaccaataaagagctctgatatgttgcaataatcaagacttaccacaaa 120
Db 124476 CTGAATAGACCAATATACAAAGTTCTGAAATTTGAGGCGCTTAATTTGACACCTTACCAACCAA 124535
QY 121 aagagtcaggagcagatgagatcacagctgaaatctctaccagagtgtaacaggaagactg 180
Db 124536 AAAAGTCCAGGACCAAGACAGATTTCACAGCCATTTCACACAGAGTTCACAAAGGAGACTG 124595
QY 181 gtaccattccctctgaaagtatacaatcaatagaaaaagaggaactctccctactcg 240
Db 124596 GTACCA-TCCCTTGAACACTATTCACAAACATATGACAAAGAGG-AATCTCTCCCAACTCA 124653
QY 241 tttaatagagccaaactatccctgatatccaaagccggcagagacacaaacaaaagag 300
Db 124654 TTTTATGAGGCCAGCATCATCTCTGATCCCAACCTGGAGAGACACAAACAAAGAA 124713
QY 301 aatttagaccataatcttgatgaacatlgatgcaaaaatctctcaataaatactgca 360
Db 124714 AATTTCAGGTCATATATCCCTGATGAATCATGATGCAAAATTCCTCATTAATATCTGGCA 124773
QY 361 aaccgaatccaagcagacatcaaaaagcttatccacatgataagtgagcttcaatcct 420
Db 124774 AACCAATCCAGACGACATCAAAACCTTATACATCAGATCAAGTTGGCTTCATCCCT 124833
QY 421 ggagat-----a 426
Db 124834 GGGATGCAAAACCTGGTTCAACAAATGTAATTAATTAATGTAATCCATCAATTAACGA 124893
QY 427 accaaagacaaaacacacatgatlatatcaatagatgcaaaaagccttgacaaaat 486
Db 124894 ACCAATTAACAAAACCAAGATATATCTCAATAGATACAGAAAAGGCTTCAACAGAAAT 124953
QY 487 caacaaccttcagttaaaaaacctcaataatagatatgtgagacatatcaca 546
Db 124954 CAAACACCCCTTCACTGCTTAAACCTCAATTAATATATCAATGTAAGTAATCTCAAA 125013
QY 547 ataataagagctatctatggaagcagcagcagcacaatactactgaatgagcaaaaatg 606
Db 125014 ATAATTAAGAGCTATTTATGACAAACCCACAGCCAAATTCATCTAGGAAGGCAAAAATC 125073
QY 607 gaagca-ttcctcttgaaaactgycacagaagaggaatgcctctctcacactctat 665
Db 125074 GAACCAAGTCCCTTGTGATTAATCTGSCACAA-----GGATGCCCTCTCTCACACACTCTAT 125128
QY 666 caatatgcttctgaaatctctgycagaggaatgaatgaagcagaggaagaaataaaggtt 725
Db 125129 CAACATAGTATTTGGAAGTTCTGCGCAGGCAATTCAGAAAGAGAGAAATTAAGGGGTAT 125188
QY 726 tcaatgagaaagaggaagtcgaatgtccctgttgtaggtgacatgtgtatccct 785
Db 125189 TCAAATAGGAAGAGAGAGAAAGTCAAAATCGCTGTCTTGCGAGATGACATGATATTTT 125248
QY 786 agaaaaccccatctctcagcccaaaaatctccttaagctgataagcaacttcagcaagt 845
Db 125249 AGAAAACCCATCGTCTCACGCCCAAAATCTCTTAAGCTGATTAAGCAAAATTCAGCAAGT 125308
QY 846 ctcaaggtacaaaatcaatgtacaaaatacaagacttctctatacccaataacagaga 905
Db 125309 CTCAGAGTACAAAATCAATGTGCAAAATCACAAGCTTCCTATACCCCAATTAACACACA 125368
QY 906 aacagagagcacaatcatgataagctcccatccaatgcttcacagagaataaata 965
Db 125369 AACAGAGAGCCCAATCTGATGATGACATCCCATTCGTATATTTGCTTCAAGGAATTAATA 125428
QY 966 cctaggaatccaaacttacaagagtgtaaggaactcttcaagagagactatacaaacact 1025
Db 125429 CTTAGGAATTAACACTTATCAAGG-----GAAGGACCTCTTAAGGACAACTATCAACACACT 125483
QY 1026 gctcaatgaaataaagaggaatacaaaaatggaagaaacatgcattgcatgagatag 1085
|||||

Query Match 22:08; Score 1005.4; DB 2; Length 149428;
Best Local Similarity 92.4%; Pred. No. 1,7e-247;
Matches 1091; Conservative 0; Mismatches 36; Indels 54; Gaps 1;

QY 1 atgataaatctcccaacatacactctcccaagactaaacccggaagattgtaactc 60
|||
Db 27266 ATGGATTAATTCCTCGACATACACCTCCCAAGACTTAAACCGGGGAAGTGTGACTCT 27207
ctgatatgacccaatacagagctctgatatgtggcaataatacagaggttaccaccaa 120
|||
Db 27206 CTGAATAGACCAATATACAGGCTCTGAAATTTGTGGCAATTAATCAATAGCTTACCAACCAA 27147
QY 121 aagagtcacagacacagatgtaattcacagctgtaattctaccagaggttacaaaggaagaaactg 180
|||
Db 27146 AAGAGTCCAGGACCAAGATGATGACAGCCAAATTTCTACAGAGGTACAAAGAGAACTG 27087
QY 181 gtaccattccctctgaaagatttaacaataatagaataaagagcaatcccttaactcg 240
|||
Db 27086 GTACCATTCCTCTGAAACATTAATCAATAGAAAAAGAGGAAATCCCTCACTCA 27027
QY 241 tttaatgagacacacatcatctctataccaaagccggcagagacacacacaaaagaag 300
|||
Db 27026 TTTTATAGAGCCAGCATCATCTCTATACAAAGCCAGGACAGACACACAAAAAAGAG 26967
QY 301 aatttagaccaatattcttgaatgaacatitgaatgaataaalcctcaataaataatctgca 360
|||
Db 26966 AATTTTAGACCAATATCTCTGATGACATGATGCAAAATCTCAATTAATAATCTGSCA 26907
QY 361 aaccgaatccagcagacacatcaaaaacttatccacatgataagtggttcaacct 420
|||
Db 26906 AACCGAATCCAGACGACATCAAAAGCTTAATCCACCATGATCAAGTGGCTTCATCCT 26847
QY 421 ggaat-----a 426
|||
Db 26846 GGGATGCAAGGCTGTCAATATATGCAATCAATTAATGATCAAGCATATAGACGA 26787
QY 427 accaaagacaaaacacatgattatctcaatagatgcagaagagccttgcacaaat 486
|||
Db 26786 ACCAAAGACAAAACACATGATTATCTCAATAGATGACAGAAAGCCTTTGACAAATT 26727
QY 487 caacaaccttcatgcttaaaaaacctcaataataatagatatgtggacatatctcaaa 546
|||
Db 26726 CAACAACCTTCATGCTTAAAACTCTCAATTAATTAAGTATGATGGACGTAATCTCAA 26667
QY 547 ataataagagctatctatggaagacagccacatcatatcgaatgggcaaaaactg 606
|||
Db 26666 ATATAAGAGCTATCTATGACAAACCCACAGCCAAATATCATAGTATGGCAAAAACTG 26607
QY 607 gaagcattcccttgaanaactgacagaagaaggaatgcccctctcacactcctatc 666
|||
Db 26606 GAAGCATTCCTTTGAAAACTGGACACAAGAGGATGCCCTCTCACACTCCTATTC 26547
QY 667 aacatagtttgaagatctggccagggcaattagcagagagaagaaataaaggttct 726
|||
Db 26546 AACATAGTGTGGAGTTCTGGCCAGGGCAATTAAGGCAAGGAATAATAAGGTAAT 26487
QY 727 caattagaaaagagagaaatcttccctgttgcaggtgacatgatactta 786
|||
Db 26486 CAATTAGGAAAGAGAGAGCAATTTGTCCCTGTTTCAGATGACATGATTTGTATATCTA 26427
QY 787 gaaaaccccatctctcagcccaaatctccttaagctgaatgaacacttcagcaagtc 846
|||
Db 26426 GAAAACCCCATTTGTCTCAGGCAAAATCTCCTTAAGCTGATTAAGCAACTTCAGCAAGTTC 26367
QY 847 tcaggtatacaaatcaatgtacaaaatacaagcatctctatacaccataacagagaa 906
|||
Db 26366 TCAGGATACAAAATCAATGTACAAAAATCAACAAATCTTATACACCAATTAACAGCAA 26307
QY 907 acagagagccaatcatgaaatgaaactcccatcaaatgtcttcaagaagaataaatac 966
|||
Db 26306 ACAGAGAGCCAAATCATGAGTGAATCCCATTCACAAATTCCTTCAAGAGAAATAAATAC 26247

QY 967 ctaggatcccaacttacaaggaatgtaaggaactcttcaaggaagaaactacaacacactg 1026
|||
Db 26246 CTAGAAATCCAACTTACAAAGGATGTGAAGACCTCTTCAAGAGAGACCTACAAACCACTG 26187
QY 1027 ctcaatgaataaagaaggtatcaacaataatggaagaacatttcatgtatagga 1086
|||
Db 26186 CTCATGAAATTAATAAGAGATACAAACAAATGAAAGAAATTCATGCTCATGGGTAGGA 26127
QY 1087 agaataataatcgtgaanaatggccatactgcccagaattat 1127
|||
Db 26126 AGAATCAATATCTGTAATAATGGCCATATCTGCCCATGTGTAAT 26086

Search completed: March 30, 2002, 13:24:16
Job time: 22938 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 09:32:04 ; Search time 5438.31 Seconds
(without alignments)
6334.866 Million cell updates/sec

Title: US-09-867-034-4

Perfect score: 3206

Sequence: 1 ttgcgctcagatgtaaac.....ttgttagatgtaaaaaaa 3206

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_estcl:*
11: gb_estc2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrl:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	401	12.5	582	10	AM939544 QV1-DT007
2	401	12.5	582	10	AM939591 QV1-DT007
3	343.6	10.7	533	11	BE871265 BE871265
4	343	10.7	517	11	BG400769 BG400769
5	325	10.1	536	10	AI791951 nm99912.Y
6	311.2	9.7	546	11	BE866879 BE866879
7	301.8	9.4	461	10	AI949096 AI949096
8	298.8	8.9	807	11	BE870652 BE870652
9	286.6	8.3	468	10	AA315623 EST187445
10	272.4	8.5	485	10	AA633302 nm58e03.s
11	260.6	8.1	477	10	AW751697 IL4-CT008
12	252.8	7.9	454	10	AI732655 nm99912.x

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	COMMENT
1	AM939544	QV1-DT0072-110200-066-a10 DT0072 Homo sapiens	AM939544	1	EST	human.	Homo sapiens	1 (bases 1 to 582)	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldstein, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	20202663	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l=et2-QV1-DT0072-110200-066-a10&t3=2000-02-11&t4=1) Seq primer: puc 18 forward

ALIGNMENTS

ALIGNMENT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	COMMENT
1	AM939544	QV1-DT0072-110200-066-a10 DT0072 Homo sapiens	AM939544	1	EST	human.	Homo sapiens	1 (bases 1 to 582)	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldstein, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	20202663	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l=et2-QV1-DT0072-110200-066-a10&t3=2000-02-11&t4=1) Seq primer: puc 18 forward

High quality sequence start: 17
High quality sequence stop: 582.
Location/Qualifiers

FEATURES

source

1. 582
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="DT0072"
/dev_stage="Adult"
/note="Organ: denis.drash. Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 180 a 154 c 93 g 155 t
ORIGIN

Query Match 12.5%; Score 401; DB 10; Length 582;

Best Local Similarity 90.6%; Pred. No. 4.5e-79;

Matches 532; Conservative 0; Mismatches 35; Indels 20; Gaps 9;

QY 390 acagctacttaccctgtccccataatagtaacacatagttccctcacaattctctaca 449
DB 16 ACAGCTACTTCACTGCTTCCCTCC--AATTAGTACACATAGTCTCCACAAATTCCTACA 73
QY 450 cctgtcccccaataatagtaacatagttccctcacaattctctacacgtcgca 509
DB 74 CCGTGTCCCTCCCAATACATACATAGTCTCCCAATTCCTATACCTACTGCTGCA 133
QY 510 gacagtgaagtaacacaaatgtcaatcagttagctacgtctgacaatacaccgttc 569
DB 134 GACAGTGAAGTCAACCAATGTAAATTTCA-TTAGCTACCTGTGACATATACCGCTTC 192
QY 570 atttccaatgatgatgattatcacaatggttcctcttgcacaagaatgaatga 629
DB 193 ATCTCCAAATGATGATTA-TCACAATGTTCTTGAACCAAGTAACATGAAA 251
QY 630 tgcctcccccacagaagaatcaatcctcagttgctccacgtcgacacgtttat 689
DB 252 TGTCCCTCCACACAGAACATCAATCATCAG-GGCTTCCCTGACCGCTTTATTG 310
QY 690 tggatgacacatgacgcttaaacagacagtgltccacgaatccttgcaagaatga 749
DB 311 GAGACGACGAC-----CCTAAACAGCACAG-GTCCGACATCTCTGCCA---AGATGAT 361
QY 750 cctgtgcagataatctgttattggttgaagcttgctataaagaattttgctgt 809
DB 362 CCTGTGCAATATATCGTTATGTGTTAACTGC---ATAATVACAAAGTTTGGCTGT 416
QY 810 gtttagaaggtatctactacaactctctacatgtaagaagaagatgttccctgag 869
DB 417 GTTTAGAAAGGATTACTACACTCTTCTACATGTAAAGAAAGAGATTTCCCTG 475
QY 870 aagattcagtgacagatatacagaacattgaccacagaagaagaacattccatg 929
DB 476 AAGATTTCACTGACAGATATCAGAAACATTGACCCACAGAGAAACATTCATGGCTAT 535
QY 930 caagaacttgcatagtgaaatctactagctgtttaaagaatgattttg 976
DB 536 CAAGACTTGCATAGTAAATTTACTAGCTGTTAAAGATGTATTGG 582

RESULT 2

AM939591 582 bp mRNA EST 30-MAY-2000
LOCUS AM939591
DEFINITION QV1-DT0072-120200-06-a10 DT0072 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM939591
VERSION AM939591.1 GI:8115050
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

source

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 582)
Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LIRC Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV1-DT0072-120200-06-a10&t3=2000-02-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 582.
Location/Qualifiers

FEATURES

source

1. 582
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="DT0072"
/dev_stage="Adult"
/note="Organ: denis.drash. Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 180 a 154 c 93 g 155 t
ORIGIN

Query Match 12.5%; Score 401; DB 10; Length 582;

Best Local Similarity 90.6%; Pred. No. 4.5e-79;

Matches 532; Conservative 0; Mismatches 35; Indels 20; Gaps 9;

QY 390 acagctacttaccctgtccccataatagtaacacatagttccctcacaattctctaca 449
DB 16 ACAGCTACTTCACTGCTTCCCTCC--AATTAGTACACATAGTCTCCACAAATTCCTACA 73
QY 450 cctgtcccccaataatagtaacatagttcctcacaattcctataactacgtcgca 509
DB 74 CCGTGTCCCTCCCAATACATACATAGTCTCCCAATTCCTATACCTACTGCTGCA 133
QY 510 gacagtgaagtaacacaaatgtcaatcagttagctacgtctgacaatacaccgttc 569
DB 134 GACAGTGAAGTCAACCAATGTAAATTTCA-TTAGCTACCTGTGACATATACCGCTTC 192
QY 570 atttccaatgatgattatcacaatggttcctcttgcacaagaatgaatga 629
DB 193 ATCTCCAAATGATGATTA-TCACAATGTTCTTGAACCAAGTAACATGAAA 251
QY 630 tgcctcccccacagaagaatcaatcctcagttgctccacgtcgacacgtttat 689
DB 252 TGTCCCTCCACACAGAACATCAATCATCAG-GGCTTCCCTGACCGCTTTATTG 310
QY 690 tggatgacacatgacgcttaaacagacagtgltccacgaatccttgcaagaatgatt 749
DB 311 GAGACGACGAC-----CCTAAACAGCACAG-GTCCGACATCTCTGCCA---AGATGAT 361

QY 750 ccctgacagataatcgtctatgttcttaagcttgcataacaagaattttgcctgt 809
|||||
Db 362 CCTGTGCATATATATGTTATGTTAAGCTGC-----ATAATACAAAGTTTGGCCCTGT 416
QY 810 gtttagaagggttactactacaacctcttcaatgttaagaagaaggatattccctgag 869
|||||
Db 417 gtttagaagggttactactacaacctcttcaatgttaagaagaaggatattccctgag 475
QY 870 aagattccagtcagacatcagacacatttgcacccagaagaagaacattccatgcctat 929
|||||
Db 476 AAGATTTTCATGACATGATCAGAAACATTTCACCCAGAGAGAAACATTTCCATGACCTAT 535
QY 930 caagacttcagtagtgaataactactagcttgccttaagaagtatttgg 976
|||||
Db 536 CAAGACTTCATAGTGAATTAATTAAGCTTGTAAAGATGATTTGG 582
RESULT 3
LOCUS BE871265 533 bp mRNA EST 20-OCT-2000
DEFINITION 601448959F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852686 5',
MRNA sequence.
ACCESSION BE871265 GI:10320041
VERSION BE871265
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 533)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LHAM9575 row: f column: 15
High quality sequence stop: 457.
Location/Qualifiers
FEATURES
source
1..533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3852686"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 157 a 142 c 111 g 122 t 1 others
ORIGIN

Query Match 10.7%; Score 343.6; DB 11; Length 533;
Best Local Similarity 86.7%; Pred. No. 2.9e-66;
Matches 436; Conservative 0; Mismatches 60; Indels 7; Gaps 5;

QY 1598 gaagcatattgaagaaggaactgttgacgaagacttcaaatcaaaactgcggt 1657
|||||
Db 1 GAAGCATATTGAAGAGAGAACTTGATGACGAAGACTTCAAAATCTAAACTGCGGT 60
QY 1658 cgaacagcttcacacatctatgagacatacgaagcgtctccctcagtcagcagattac 1717
|||||
Db 61 CG-ACAGGCTTCACCAATCT-TGAGAGAGAGAGGAGCGCTCTTCCCTAAGGTCAGGATAC 118
QY 1718 ggcctcacaagagacgcctagatgcaaaaatcccgtagttccaagacacagcagatgcc 1777

Db 119 GACCTGCC---AGAGACAGCAGATGCAGAAATCCCGTAGTTTCAGACAGACAGCATG-C 174
|||||
QY 1778 ccggcctgactattagaatccatcagaatgttgaaccgccatgcccccaacatalg 1837
|||||
Db 175 CCGCGCTGACATATTAGATCATAGAAATGTGGAACCMCGCATGCGCCCAACCATATG 234
QY 1838 tacatctattatctcagcagtggtttgacagaagctcatgagaagtgaagacacagct 1897
|||||
Db 235 TACATATCTATTATCTAGACAGTGTGACAGAGCTGATGAGAGAGAGACACACAGCT 294
QY 1898 aaagactcggcctccgggaagttctctccatcctagacatactgcagctccatctg 1957
Db 295 AAAGACTCTGGCGCTCGGAGAGTTTCTTCTTCATCATGACATACAGCAGCTCATCTG 354
QY 1958 caatgcaacgcttgcgaatgtcttgcgaagacatccagcgtcctgtctaaataaga 2017
|||||
Db 355 CAATGCAACGTTGTGCAATGTCTGCAAAACGACATCCAGCTCACTTGGT-AAATTAAGA 413
QY 2018 atctatgacattacaatgtagctgtagtctatgacgctgctcagagaggttggttt 2077
|||||
Db 414 ATCTATGACATTTACATATGACTGCATGCTTATTAAGCGCTTGACAGCAGCGGTGCTC 473
QY 2078 ctccaatcagtaacaagaagtaactg 2100
|||||
Db 474 CTCCTCACTCAAGCTACAAACAG 496
RESULT 4
LOCUS BG400769 517 bp mRNA EST 12-MAR-2001
DEFINITION 602464172P1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4592489 5',
MRNA sequence.
ACCESSION BG400769
VERSION BG400769 GI:13294217
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 517)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@email.nih.gov
Tissue Procurement: Clontech Laboratories, Inc.
CDNA Library Preparation: Clontech Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LHCML130 row: o column: 18
High quality sequence stop: 517.
Location/Qualifiers
FEATURES
source
1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4592489"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pDNR-LIB (clontech); Site_1:
SfiI (ggcggcctcgcc); Site_2: SfiI (ggcctatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor sequence:
5'-ATCTAGAGCGCCGAGCGCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."
BASE COUNT 166 a 118 c 93 g 140 t

ORIGIN

Query Match 10.7%: Score 343; DB 11; Length 517;
 Best Local Similarity 90.6%: Pred. No. 3.9e-66;
 Matches 481; Conservative 0; Mismatches 30; Indels 20; Gaps 10;

QY 521 aaccacaatgtaattcagttagctacccctgacataaccgcgtccatcccaatg 580
 |||||
 Db 1 AACCAACAATGTAATTCGA-TTAGCTACCTCTGACATATACCGGCTTCATCTCCAAATG 59
 QY 581 atgagtaattcacaatggttcctcttgaaacaaagttaacatgtaaatgtccacc 640
 |||||
 Db 60 ATGAGATTAA-TCACAATGTTCTCTTGAAACACAAGTAACATGAATGTCCCCACC 118
 QY 641 acagaagacaatcaccatgagctccaccctgagcagccgtttatttggaagaccat 700
 |||||
 Db 119 ACAGAAAGACATATATCATCAG-GGCTCCCACTGGCACCCTTTATYGGAGACACGA- 176
 QY 701 gcaagcctaacaacagcagctgtccagcaatccttgccaagaatgatacccccgtgacga 760
 |||||
 Db 177 -CCCTTAACAGCAGAG-GTCCAGCAATCCTTGCCA--AGATGATCCCTGTGCAGA 228
 QY 761 taattcgtaattgttgtaagctgtcataaacaagttttgctgtgtttaagaagg 820
 |||||
 Db 229 TAATTCGTATGTGTAAAGCTGC---ATAATACAGTTTTCCTGTGTAAAGAGG 283
 QY 821 taattactaacctctctacatgttaagaagaagaagattcccttgagaagaatttcagt 880
 |||||
 Db 284 TATTACTACACTCTCTCTACATGATGAAGAAAGAAAGTATTCCTCG-GAAGATTTCAGT 342
 QY 881 gacagatcacagaacaattgaccagaagaagaacattccatgagccataaacttgca 940
 |||||
 Db 343 GACAGATATCAGAAACATTTGACCCAGAGAGAAACATTCATGAGCCATACAACTTGCA 402
 QY 941 tagtgaattactacagcttgttaagaatgtatttgacacatcgtttatggaagactgt 1000
 |||||
 Db 403 TAGTGAATTTACTAGCTGTGTTAAAGATGTATTTGGCACATGCTTTATGACAGACTGT 462
 QY 1001 aactctactgtaagacacatcctctgacacaagattctaaatgcagct 1051
 |||||
 Db 463 AATTCTTACTGTAA-GCACATCTCTGTCAACCAAGA-TCCTAATAGCTGTGT 511

RESULT 5
 LOCUS A1791951 536 bp mRNA EST 13-DEC-1999
 DEFINITION nm99g12.y5 NCI-CGAP_C09 Homo sapiens CDNA clone IMAGE:1076422 5',
 mRNA sequence.
 ACCESSION A1791951
 VERSION A1791951.1 GI:5339667
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 536)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Other ESTs: nm99g12.x5
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.
 , Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
 Original clone citation: see original entry for original citation
 information
 This 5' resequenced clone has no previous 5' data to verify this
 new read against
 Insert Length: 581 Std Error: 0.00
 Seq primer: -40RP from Glibco
 High quality sequence stop: 453.
 Location/Qualifiers
 1. 536
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1076422"
 /clone_lib="NCI-CGAP_C09"
 /tissue_type="colon tumor RER+"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT7n3D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 RER+ colon tumor, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT7n3
 vector. Library is not normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo (Soares4
)."

BASE COUNT 149 a 118 c 131 g 136 t 2 others
 ORIGIN
 Query Match 10.1%: Score 325; DB 10; Length 536;
 Best Local Similarity 89.3%: Pred. No. 4.1e-62;
 Matches 499; Conservative 0; Mismatches 37; Indels 23; Gaps 13;
 QY 2627 aaaaacatgaccctgtgtaagaagaagaagaagaactggtggaagatcaatla 2686
 |||||
 Db 1 AAAAACATGACCTGTGTAAGAAAGAAAGAGCAAAAGAACTGGTGGGAGATCAATTA 60
 QY 2687 gagagagagcactggatggatccactgcttccttaagtlccctccatcagcaagaag 2746
 |||||
 Db 61 GAGAGGAGGACACCTGGGATCCACCTTC-TTCCTTAGGTCCCTCCCTCCAT-CAGCAAAAG 118
 QY 2747 agcactctctaagtcatactgctccctccgaagactgctgaggaagaagtttaaaacaaa 2806
 |||||
 Db 119 AGCAGTCTCTTAA-TCATGCCCTCCGAAAGACTGGGAGAGGTTT-AAAAACAAAA 176
 QY 2807 aatccagagtaagaagccttaagtgatcagtttgaaaattggaacaaactgctcggc 2866
 |||||
 Db 177 AATCCAGAGAGT-AAAGACCTTAGG---TCAGTTTGAATTTGAGACAAACT--CTCGGC 230
 QY 2867 aaaggtgcacaagacgagactgttgctcagaagatccagccgctcagcctcgaggtgt 2926
 |||||
 Db 231 AAAGCGTGC---GAGAGGAGGAGTTGTCTCAGAGAGT-CCAGCCCTCCAGCCTCGGGGTGT 286
 QY 2927 aaggtctcagtgatgtgacatgaggggctcagcctcttcctgtaaccagggctcagctg 2986
 |||||
 Db 287 -AGCTTCTTAGGCTGCATGTGGGCTCAGCCCTTCTGTGTA-CAGAGGCTCACCTG 344
 QY 2987 tggccacaac 3046
 |||||
 Db 345 TGGCCACCAACACACACACACACACACACACACACACACAAATGGGGCAACACATC 404
 QY 3047 ccacgtacccaagccttaacacaaatgtaattagtgctccctttatttctaataagcct 3106
 |||||
 Db 405 CAGTACAAAGC-----TTTAAACAATGTATTTAGTGTCTTTTATTTCTTAAT-GCCTT 457
 QY 3107 gtccctctaagaattattatttatttattattatttatttatttatttatttatttattt 3166
 |||||
 Db 458 GTCCCTTTAATAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 517
 QY 3167 ggtcaatgcaataagtcgc 3185
 |||||
 Db 518 GGTATATGCAATANAAGTCC 536

RESULT	6
LOCUS	BE866879 546 bp mRNA EST 20-OCT-2000
DEFINITION	601442694.F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847054 5' , mRNA sequence.
ACCESSION	BE866879
VERSION	BE866879.1 GI:10315655
SOURCE	EST.
ORGANISM	human. Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 546)
TITLE	NIH-MGC http://mgc.nci.nih.gov//.
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: csapbs@remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LHAM9560 row: k column: 23 High quality sequence stop: 540. Location/Qualifiers 1..546 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3847054" /feature_type="adenocarcinoma" /lab_host="DHI0B (phage-resistant)" /note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."
FEATURES	
Source	
BASE COUNT	155 a 126 c 131 g 134 t
ORIGIN	
Query Match	9.7%; Score 311.2; DB 11; Length 546;
Best Local Similarity	87.9%; Pred. No. 4,9e-59;
Matches 503; Conservative	0; Mismatches 49; Indels 26; Gaps 14.
OY	1154 aagtagctcaagaacatttctaactatgatgtgaccctgcgtgtgttgatgtttgag 1213
Db	1 AAGTGCTCAAGCACC -TTTCFAACTATGATTGACCC-----TTGGCGGTATTT 53
OY	1214 ggcctggagacaaggacttcgtgatgactgcctcaaatggtttgaagatgcatgtgcaaa 1273
Db	54 ATGGCTGTAAACGACTG--CGGATGACCTGCCCTCAAT-GGTTTACCATGCGA--TTGCAA 108
OY	1274 tgctaacctcgaaaagcccttaaccaagagaccttcctgcgtgtctccaagtccaagt 1333
Db	109 ATCTGAAGCTGCAAGAGCCCTTAACCCACAGAGCCCTTTCTCGCTTCAGCTCA-AAT 167
OY	1334 gtccctgatgcctcgcaagcagcacgaagaagcgaatgcttaataagaagaagtgtgagtc 1393
Db	168 GTCCGATGAGCCCTGCAAGCAGCACAGCAACGATGCTTAATAAGAAGAGTGGTGGGCC 227
OY	1394 cccttcagatgttgtgtgtgtgtgccgtgtaccagagaagatgtcataaggaactgcaaa 1453
Db	228 CCTGAGTGTGCTGCTG---CCCGGCTTACAGGAAATCTTAATGGAACTGCCAAA 282
OY	1454 agtgcgcatltyggctacagcttgacttcgactgtaaagacaatatccaagtccact 1513
Db	283 AGTGTGCAATTT-GGCTACAGTGTGAGCTCGACTGTAAAGACAATTTTCAGTGAATCCCTCACT 341
OY	1514 tatttgyggagcaacatcgctgcgcatltgtcatctcagaatgataatgtcatgtattgtca 1573

Db	342	--ATTGTGGGGCCACCATTCGCTGCATTTGTCATTTCTCAGCATGATGAATTTGCATTGATTGTCA	399
Oy	1574	ctagcaagatcacaatacaaaacgaagcatatltgaagaagaaacttgattgacgaag	1633
Db	400	C-AGCAAGATCAAAATACAAAA-CGAAGCATATTGAGAA-GAGAACTTGATTGACGAAG	456
Oy	1634	acttcaaatcaaatcaacgctgcagcagcttcacacatctatgagatcaagag	1693
Db	457	ACTTTCAAAATTAATAACGCGGTGC-ACAGGCTTACCAATCT-TGAGACAGAAGGAG	514
Oy	1694	cgctccctcagtcagtcagatcacgctcca	1725
Db	515	CGCTTTCTAGTCAGCATGAATACGCTCCA	546
RESULT	7		
AI949096			
LOCUS			
DEFINITION	AI949096	461 bp	mRNA
ACCESSION	AI949096		EST
VERSION	AI949096.1	GI:5741406	08-MAR-2000
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
Source			
Query Match	9.4%	Score 301.8	DB 10; Length 461;
Best Local Similarity	98.1%	Pred. No. 6, 2e-57;	
Matches 358; Conservative	0;	Mismatches 2;	Indels 5; Gaps 5
Oy	829	caacctctacatgtaaaagaaaggtatcccttgagaagatcttcagtacagat	888
Db	22	CAACCTCTTACATGTAAAGAAAGGAAGATATTCCTG-GAAGATTTCAGTACAGTAT	80
Oy	889	cgaaacatctgaaccagaagaacatctccatggcctatcaagaacttgacatagtaaa	948

```

Db      81 CAGAAACATTTCACCCAGAGAGAAACATTTCATGGCTATCAAGACTTGCAATAGTGA 140
Oy      949 tcaacagctctgttaaagatgtaattgacacatctgttaagacagactgtatctta 1008
Db      141 TTACTAGCTGTTTAAAGATGATTTGGACATCTGTTATGACAGACTGTAACTCTTA 200
Oy      1009 ctgaagacacatctctgacccaagatctgaatgctgcttgatgaagaatttgtt 1068
Db      201 CTGTAA-GCACATCTCTGTCCACCAAGATTCGAAATGCGTGTGATGACAAAGTTTGT 259
Oy      1069 aatgtaacatgttaacaattttggcagaaccacagaagtgacaatgaagaactgtgact 1128
Db      260 AATGTAACTATGTAAACATTTTGGCAGAACCCAAAGTGACATGAGAAGACTGTGACT 319
Oy      1129 ggaagaataataaagacattataagtaagtcgaagcaacttcttaactatgttg 1188
Db      320 -GAGAAATTAATAAAGCAA-TTAGAAGTAGCTCAAGCAAC-TTCTAAACTATGATTTG 376
Oy      1189 accct 1193
Db      377 ACCCT 381

RESULT 8
LOCUS   BE870652      807 bp      mRNA      EST      20-OCT-2000
DEFINITION 601447779F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3851730 5',
VERSION  BE870652
KEYWORDS BE870652.1 GI:10319428
SOURCE  EST.
ORGANISM human.
          Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 807)
          NIH-MGC http://mgi.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-tr@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LHM9572 row: n column: 19
          High quality sequence stop: 652.
          Location/Qualifiers
            1..807
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:3851730"
              /clone_id="NIH_MGC_65"
              /tissue_type="adenocarcinoma"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI;
              Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
              Average insert size 1.8 Kb. Library constructed by Life
              Technologies."
BASE COUNT      234 a      193 c      214 g      166 t
ORIGIN
Query Match      9.3%; Score 298.8; DB 11; Length 807;
Best Local Similarity 79.9%; Pred. No. 2.8e-56;
Matches 638; Conservative 0; Mismatches 102; Indels 58; Gaps 22;
Oy      2261 cgcatacactgagaggtccgagggagatgagaagaataccacacatcttcataagg 2320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```

```

Db      57 CGCATCACTGTGAGGCTCCGAGGG--ATGAGAGGGATACCACCCTTTCAAGGCT 114
Oy      2321 tcacaagctacacactcgtgtgacagatcagaatagggacacctgtctctatccctccat 2380
Db      115 CACAAGCTCAGCTTC--TGACAAAGTCAGAAATAGGGACA--CTGCTTCTATCCCTCCAT 169
Oy      2381 ggagagatcttggccaaccccccttlttgaaaacgaagcccccagaagcttggcaac 2440
Db      170 GGA-GAGATTTCTGGCAACCTTTG-----AACAGCCAGAGCTT-GCAAC 211
Oy      2441 ctgacctcaacccaagaagactgaaagagacatcttcttaagcttcttggagagcg 2500
Db      212 CTAGCCTC-ACCCAGAAGACTGGAAA-GAGACATATCTC-TGAGCTTTTTCAGGAGCG 268
Oy      2501 tgccttgggaatccagaagcttcttgatgtaattagaagcgctgactatataatgt 2560
Db      269 TGCC-TGGGATCCAGGAAAC-TTTTGATGCTATTAAGAAGCGCTGGACTAAAAATGTC 326
Oy      2561 ccatctatgggttctaatacaglttltgaacatgctagaagcagaacgaggccaga 2620
Db      327 AC---TATGGGGTGACACTTACAGTTTGGAA-ATGCTAGCAGCAAGAGGGCGCAGAGA 381
Oy      2621 gagtaaaaaaacatgacctggtgtagaagagaagcagaagaaactggtggagagat 2680
Db      382 G---TAAAAAACATGACCTGCTAGAGAGAGAGCAAGCAAGCAACTGGGTGGAGAGAT 438
Oy      2681 caatagagagagagacactggtgacacactgcttcccttaagtcctccctccatgag 2740
Db      439 CATTTAGAGAGAGGACACCTGGGATCCACCTTC-TTCCTTAGGTCCCTCCTCCAT-CAG 496
Oy      2741 caaagagacacttcttaagtcacatgcccctccgaagacttggtagaagagtttaaaa 2800
Db      497 CAAGAGAGACCTTCTTAA-TCATGCCCTCCCAACACTGCTGGGAGAGAGGTTTAAACA 555
Oy      2801 acaaaaaatcccaagagctaaagaccttaggtgcagtttgaagaattggagaacactgt 2860
Db      556 ACAAATAATCCAGAGT-AAGAGCCTTAGG---TCAGTTTGAATCGGACACAAC---T 608
Oy      2861 ctgtgcaaaaggtgtgcaaaagcgaagctgtgtgctcaagagatccacgctccagctcg 2920
Db      609 GTCGTGCAAAAGGTGCGAGAGCGGAGCTTGCTCAGAGAGT-CCAGCCCGTACGCGCTCG 667
Oy      2921 ggggtgaaggtctgtgaaggtgtgcatggtggggccttaagccttctgtgtgacccgagct 2980
Db      668 GGGTGTAAAGTCTCTAGGTGTGTCATGGGGGCTCAGCCTTCTGTGATGACCCGAGGTC 727
Oy      2981 cagctgtgcccacacacacacacacacacacacacacacacacacacacacacacac 3040
Db      728 AGTTGTGGCACAAAAAACA-----ACACAAAACACACACACATGSGGCACAC 778
Oy      3041 ccacatccacgttaacca 3058
Db      779 CTCATATCAAGCTTACAA 796

RESULT 9
LOCUS   AA315623      468 bp      mRNA      EST      19-APR-1997
DEFINITION EST187445 Colon carcinoma (HCC) cell line II Homo sapiens cDNA 5',
          end similar to similar to cell surface antigen 114/AlO, mRNA
          sequence.
ACCESSION AA315623
VERSION   AA315623
KEYWORDS  EST.
SOURCE    human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 468)
          Adams,M.D., Kerlavag,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
          ,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D., White
          ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Whi,C., Clayton,R.A.,
          Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald

```

L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodok, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S.Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, X.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
 JOURNAL Nature 377 (6547 Suppl.), 3-174 (1995)
 MEDLINE 96026280
 COMMENT Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tcdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES
 source Location/Qualifiers
 1..468
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):111404"
 /db_xref="taxon:9606"
 /clone_lib="Colon carcinoma (HCC) cell line II"
 /tissue_type="colon"
 /cell_line="KM12C"
 /note="Organ: colon; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 153 a 90 c 95 g 130 t
 ORIGIN

Query Match 8.9%; Score 286.6; DB 10; Length 468;
 Best Local Similarity 91.1%; Pred. NO. 1.5e-53;
 Matches 440; Conservative 0; Mismatches 24; Indels 19; Gaps 12;

QY 829 caactctctacatgtaagaagaagatattccctggaagaagattcgaatgacagat 888
 |||||||
 Db 4 CAACCTCTTACATGTAGAGAAAGGATATTCCTCGG-GAAGATTTCAGTACAGTAT 62
 |||||||

QY 889 cagaacaattgaccagaagaagaacattccatgacctatcaagaacttgatgtaaa 948
 |||||||
 Db 63 CAGAAATTTGACCCGAGAGAGAAACATTCATGACCTATCAAGACTTGATGTGAAA 122
 |||||||

QY 949 ttactaactggttaagaagatattggcaccatctgtttatgacagacttaattcta 1008
 |||||||
 Db 123 TTACTACTCTGTAAAGATGATTTGGACATCTGTTATGAGACAGCTGTAATTTCTTA 182
 |||||||

QY 1009 ctgtaagcacatctcgtcaccaagaattcgaatcgcttctgatacaagaattggt 1068
 |||||||
 Db 183 CTGTAA-GCACATCTCTGTACCAAGA-TCTGAATGCGTTC-TGATGACAG-TTTTGTT 238
 |||||||

QY 1069 aatgtaacaatgatacaatttggcagaacaacaagtgcacaaatgagaagaactgtgact 1128
 |||||||
 Db 239 AATGTAAATAGTAAACAATTTTGGCAGAAACCAAGTGACATGAGAAACAGTGTGACT 298
 |||||||

QY 1129 gaggagaattaagaacaattataagtagctcaagaacaatttctaataatgattgg 1168
 |||||||
 Db 299 -GAGAAATTAATAACCA-TTAGAGTAGCTCAAGCAAC-TTCTTAAGTATGATTTG 355
 |||||||

QY 1189 accctgctggttgatgattgagagctgggaaccaagactgctgatgacctca 1248
 |||||||

|||||
 Db 356 ACCG-----TTCCGTGATGATATTATGCGGTAAACGACAGCTG--CGGATACAGTCCCA 407
 |||||||

QY 1249 atggattgacatgcgatgataatgctgacatgcacaaaggcctaaccacagagccct 1308
 |||||||

Db 408 AT-GGTTTGGATGATGGA--TTGCAATCTGACCTGCAAGGCTTAACCCAGAGACCTT 464
 |||||||

QY 1309 tct 1311
 |||

Db 465 TCT 467

RESULT 10
 AA633302/c 485 bp mRNA EST 21-OCT-1997
 LOCUS n958e03.s1 NCI-CGAP_C09 Homo sapiens cDNA clone IMAGE:1148092 3',
 DEFINITION mRNA sequence.
 ACCESSION AA633302
 VERSION AA633302
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 485)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 452.

FEATURES
 source Location/Qualifiers
 1..485
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="1148092"
 /clone_lib="NCI-CGAP_C09"
 /tissue_type="colon tumor RER+"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pRT3D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 RER+ colon tumor, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pRT3
 vector. Library is not normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo (Soares4
)"

BASE COUNT 135 a 113 c 112 g 125 t
 ORIGIN

Query Match 8.5%; Score 272.4; DB 10; Length 485;
 Best Local Similarity 88.3%; Pred. NO. 2.3e-50;
 Matches 447; Conservative 0; Mismatches 36; Indels 23; Gaps 13;

QY 2692 gaggacactggatccacactcgttccctaggtcccccctccatgcagcaagagcac 2751
 |||||||

Db 485 GAGGCACTGGGATCCACACTTC-TTCCTTAGTCCCTCTCCAT-CAGCAAGAGAGAC 428
 |||||||

QY 2752 ttctctaagcatagccctccgaagaactgctgggagaagtttaaaaaaacaatcc 2811
 |||||||

Db 427 TTCTCTAA-TGATGCCCTCCGAGAGACTGCTGGAGAGAGTTT-AAAAACAAAATCC 370
 |||||||

OY 2812 agagagtaagagccttaggttagttagttagaataatggagagacaaactgtcttggcaag 2871
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 369 AGGAGT-AGGAGCCTTAGG---TCAGTTGAATGTGAGCAAACT--GCTGCGCAAGG 316
OY 2872 gggcaagagcgagagcttctgctcagagagctcccaagccctccgaggtgttagt 2931
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 315 GTGC---GAGAGGAGCCTTGCTCAGGAGT--CCAGCCGTCAGCCCTGGGGTGT-AGGT 261
OY 2932 cctctggaggtgctccatgggggctcagcctctctgtgtgacccgagctagctgtgccc 2991
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 260 TTCTGAGGTGTGCTATTTGGGGCTTCAGCCTTCTCTGTGA-CAGAGGCTCAGCTGTGGCC 202
OY 2992 accaac 3051
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 201 ACCAACAACACACACACACACACACACACACACACACAAATGGGGCAACACACATCCAGTA 142
OY 3052 taaccaagcttaacacaaatgtatagtgtcccttttatttcaataagccctgtcct 3111
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 141 CAGC-----TTTAAATATGTTATAGTGTCTTTTATTTCTTAAT-GCCTTGCTT 89
OY 3112 cttaaaagtatttatttcttatttatttcttctgactgttaattgtaattgtaa 3171
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 88 CTTAAAGTATTATTATTGTTATTTATTTATTTCTTCTGACTGTATATGTGATGTATA 29
OY 3172 tgcataaagtgcttctgttagatg 3197
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 28 TGCATTAAGTGCCTTTGTAGATGG 3

RESULT 11

AW751697 477 bp mRNA EST 28-APR-2000
LOCUS AW751697
DEFINITION IL4-CT0081-130999-001-C05 CT0081 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW751697
VERSION AW751697.1 GI:7666629
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 477)
Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

CONTACT: Simpson A.J.G.
LABORATORY OF Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/ILICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL4&t2=IL4-CT0081-
130999-001-C05&t3=1999-09-13&t4=1)
Seq primer: puc 18 forward.
Location/Qualifiers

FEATURES

SOURCE

1.477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0081"
/dev_stage="Adult"
/note="Organ: colon; Vector: pUC18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived

from ORFESTS PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 139 a 107 c 114 g 117 t
ORIGIN

Query Match 8.1%; Score 260.6; DB 10; Length 477;

Best local similarity 86.2%; Pred. No. 9,7e-48;

Matches: 432; Conservative 0; Mismatches 44; Indels 25; Gaps 12;

OY 1068 taatgaacaaatagtaacaaatttggcagaacacacaaagtgacaatgaagaagctgtgac 1127
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 2 TGATGTAAACAAATAGTGAATTTGGCAGAAACACAAAGTACAAATGAGAAAGCTGGAC 61

OY 1128 tggagaagaatataagaacattataagaagctcaagaacatttctaactatgattg 1187
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62 T-GAGAAATCAATTAAGCAATT-AGAAGTGTCTCAAGCAAC-TTCTTAACATATGATTT 118

OY 1188 gaccctgcgtgtgattgattgaagggctgggaaccaagactgctgtgactgctc 1247
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 119 GACCC-----TTGGAGTGAATATTTATGCTGTAAACCAACTG--CGATGACTGCTC 170

OY 1248 aatgggttagacatgcatgtgcaaatgctgacctgcaaaagccttaacacagagcct 1307
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 171 AAT-GGTTTAGCATGCACTG---CAATCTGACCTGCAAGAGGCTTAACCCACAGAGCCCT 226

OY 1308 ttctgtgcttcccaagctcagaagtgctcgtgctgctgcaagcagacagaagaaggaat 1367
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 227 TTCTGCTGCTTCCAGCTCA-AGTGTCTGATGCTTGAACGACAGCAAGC-AAT 284

OY 1368 gcttaataagaagatggtgtggggtccctgcagtgctgtgctgtgctgtctacca 1427
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 285 GCTTAATTAAGAAAGAGTGTGGGGCCCTGAGTGTGCTCGTG-----CCCGGCTACCA 339

OY 1428 ggaagatgcatatggaactgccaagaagtgtcatattggctacagtgtgactgtgta 1487
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 340 GGAAGATGCTAATGGAAGTCCCAAAAGTGTGCAATT-GGCTACAGTGCAGTCACTGTA 398

OY 1488 aggaacaatttcagctgactcactcatttggggcacatcgtcgtgacttcttc 1547
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 399 AGAGCAAAATTCAGCTGATCTCACT--ATTGTGGGCAACCATCGCATGTGCAATTCT 456

OY 1548 cagcatgataatgcatgtat 1568
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 457 CAGCATGATATTCATTTGAT 477

RESULT 12

AI732655 454 bp mRNA EST 13-DEC-1999
LOCUS AI732655/c
DEFINITION nm99412.x5 NCI-CGAP_C09 Homo sapiens cDNA clone IMAGE:1076422 3',

mRNA sequence.
ACCESSION AI732655

VERSION AI732655.1 GI:5053768
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 454)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.

JOURNAL

COMMENT

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
 Original clone citation: National Cancer Institute, Cancer Genome
 Anatomy Project (CGAP), Tumor Gene Index
 This read has been verified (found to hit its original self in the
 correct orientation)

Insert Length: 581 Std Error: 0.00
 Seq primer: -400P from Gibco.

FEATURES

source

1..454
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI_CGAP_Cg9"
 /tissue_type="colon tumor RER+"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 RER+ colon tumor, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library is not normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo (Soares4
)."

BASE COUNT 125 a 102 c 104 g 123 t
 ORIGIN

Query Match 7.9%; Score 252.8; DB 10; Length 454;
 Best Local Similarity 87.6%; Pred. No. 5.3e-46;
 Matches 417; Conservative 0; Mismatches 37; Indels 22; Gaps 12;

Oy 2717 ccttaagtccttcctcctcatgcaagaagagcactctcctcaatgacccctccgaag 2776
 |||||||
 Db 454 ccttaagtccttcctcctcatgcaagaagagcactctcctcaatgacccctccgaag 397
 |||||||
 Oy 2777 actgctggaagaaggtttaaaaaaacaacaaatccagagtaagaagccttagagttcag 2836
 |||||||
 Db 396 actgctggaagaaggtttaaaaaaacaacaaatccagagtaagaagccttagagttcag 342
 |||||||
 Oy 2837 ttggaatctggagacaactgtcttgcaaaaggtgccaaagcgagctgtgtctc 2896
 |||||||
 Db 341 gtttgaaattggagacaaact--gtctggcgaagcgttc--gaagagcgaccttgctgtc 287
 |||||||
 Oy 2897 aggaagtcacagcgctcagcagcttgaggtgtgaagtcctcgtgagtggtgcagtgagcgctc 2956
 |||||||
 Db 286 aggaagt--cagacctctcagcctcgagggtgt--aggtttctgaggtgtgctcattggggcctc 229
 |||||||
 Oy 2957 agcctctctgtgacccgaagctcagctgtggaaccaacaacaacaacaaca 3016
 |||||||
 Db 228 agcctctctgtgacccgaagctcagctgtggaaccaacaacaacaacaacaaca 170
 |||||||
 Oy 3017 accacaacaacaatgggggcaaccacatccacgttaaccaagctttaacaacaatgtta 3076
 |||||||
 Db 169 accacaacaacaatgggggcaaccacatccacgttaaccaagctttaacaacaatgtta 116
 |||||||
 Oy 3077 tttagtgccttttatttcttaataagcctgtctcttaaaagtattttttgtatt 3136
 |||||||
 Db 115 tttagtgccttttatttcttaaat--gccttgctccttttaaaagtattttttgtatt 57
 |||||||
 Oy 3137 ataatctgtcttctgactgtaaatgtgaatgtaatgcaataaagtgcccttgta 3192
 |||||||
 Db 56 attattgttcttgcactgttaattgtgaatgtaatgcaataaagtgcccttgta 1

RESULT 13
 A1573096/c A1573096 449 bp mRNA EST 14-MAY-1999
 LOCUS A1573096

DEFINITION trf9e02.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2223578 3',
 mRNA sequence.
 accession A1573096
 version A1573096.1 GI:4536470
 keywords EST.
 source human.
 organism Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 449)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapb-remail.nih.gov
 Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 2580 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 410
 POLYA-No.

FEATURES location/qualifiers

source

1..449
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:2223578"
 /clone_lib="NCI CGAP_Pan1"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"

BASE COUNT 125 a 104 c 99 g 121 t
 ORIGIN

Query Match 7.8%; Score 250.4; DB 10; Length 449;
 Best Local Similarity 87.6%; Pred. No. 1.8e-45;
 Matches 403; Conservative 0; Mismatches 36; Indels 21; Gaps 11;

Oy 2738 cagcaagagagcactctcttaagtcagccctccgaagcgtgctggagaagtttaa 2797
 |||||||
 Db 441 cagcaagagagcactctcttaagtcagccctccgaagcgtgctggagaagtttaa 384
 |||||||
 Oy 2798 aaaaacaaaacccagagtaagaagccttagagtcagtttgaaaattggagacaact 2857
 |||||||
 Db 383 aaaaacaaaacccagagtaagaagccttagagtcagtttgaaaattggagacaact 328
 |||||||
 Oy 2858 tctctggcaagaggtgcaagagcgagctgtgtctcgaagagtcacagccgtcagcc 2917
 |||||||
 Db 327 --gtctggcaagaggtgc--gagagagagctgtgtctcagagag--cagccgcccagcc 274
 |||||||
 Oy 2918 tcgggtgttaagtcctcgtgaagtgcaatggggcctcagcctctctcgtgacccag 2977
 |||||||
 Db 273 tcgggtgtgt--aggtttctgaggtgtgcatatgggctctcagcctctcgtgacccag 216
 |||||||
 Oy 2978 gctcagcgtgtgcaacaacaacaacaacaacaacaacaacaacaacaatgggggc 3037
 |||||||
 Db 215 gctcagcgtgtgcaacaacaacaacaacaacaacaacaacaacaacaatgggggc 156
 |||||||
 Oy 3038 aaccacatccaagtaaccaagcttaacaacaatgtattagtgccctttattct 3097
 |||||||
 Db 155 aaccacatccaagtaaccaagcttaacaacaatgtattagtgccctttttttttct 102
 |||||||
 Oy 3098 aatagccctgtcctcttaaaagtatttattgttattatttcttctgtactgtta 3157
 |||||||
 Db 101 aat--gccttgctccttttaaaagtattttattgtattattgtgtgactgtta 43

```

QY 3158 attggaatgtaatgcaataaagtcctgttagatgg 3197
Db 42 ATTGTGAATGTAATGCAATAAAGTCCTTTGTAGATGG 3

RESULT 14
AM363038/c 531 bp mRNA EST 04-FEB-2000
LOCUS RC0-CT0301-271199-031-G12 CT0301 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM363038
ACCESSION AM363038
VERSION AM363038.1 GI:6867688
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 531)
HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL:
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC0&c2=RC0-CT0301-
271199-031-G12&t3=1999-11-27&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 14
location/Qualifiers
1..531
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0301"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site.1: SmaI; Site.2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 112 a 155 c 117 g 147 t
ORIGIN
Query Match 7.7%; Score 246.6; DB 10; Length 531;
Best Local Similarity 87.0%; Pred. No. 1.3e-44;
Matches 481; Conservative 0; Mismatches 44; Indels 28; Gaps 18;

QY 2450 acccaagaagactggaagaagacacatcttcaagcttttcaagaagcgctgacctgag 2509
Db 525 ACCCAAGAAGACTGGAAA-GAGACATATC-TCTACAGCTTTTCAGAGAGCGGTGCC-TGGG 469

QY 2510 aatcagaagaagcttttgaagtcataatagaagcgctgagcctaataatagtcacatcag 2569
Db 468 AATCAGAGAAC-TTTTGTGCTAATTAGAGAGCCCTGAGCTAAATGTCAC----TAT 414

QY 2570 gggcttaatcacagctttgaacatgctagagcagaacgggagccagaagataaaa 2629
Db 413 GGGGTGCACTCTACAGCTTTTGAA-ATGCTAGAGAGCAGAGGCGCAGAGAG----TAAA 358

QY 2630 aacatgacctggtlagaagaagaagagcaaaactggtggtggaagatcaatagag 2689
Db 357 AACATGACCTGTGTAGAGAGAGAGCAAGCAAACTGGGTGGGAGAGATCAATTAGAG 298

QY 2690 aggaagacctggtgagatccaccttgccttagtgcctcctccatgcagcaagaagagc 2749

```

```

Db 297 AGAGAGCACTGGGAATCCACCTTC-TTCCTTAGGTCCCTCCCTCCAT-CAGCAAGAGC 240
QY 2750 actctctaagtcaccccccacgaagactgctggaagaaggtttaaanaaaaaaat 2809
Db 239 ACTTCTCTAA-TATGCGCTCCCGAAGACTGCTGGAGAGAGTTT-AAATAAATAAAT 182
QY 2810 ccagaagtaaaagaccttaaggtcagtttgaanaattggaacaaactgtcttgcaaa 2869
Db 181 CCAGAGAGT-AAGAGCCTTAGC---TCAGTTTGAATTGGAGACAACCT-GTCTGGCAAA 128
QY 2870 ggggtccaaagcgcgagctgtgttcaggaagatcccaagccgtcagcctgggtgtaag 2929
Db 127 GGGGTGC--GAGAGGAGAGCTTGTGCTCAGAGAT-CCAGCCGCCACGCTTCGGGGTGT-AG 73
QY 2930 gctctgaagctgctcatggggggcccaagcctctcgtgagccgaggtcagctgtg 2989
Db 72 GTTCTGAGGTGTGCAATTTGGGCTCAGCCTTCTGTGCTGA-CAGAGGCTCAGCTGTGG 14
QY 2990 ccaccaacacaca 3002
Db 13 CCGCAACACTCA 1

RESULT 15
AA581944 431 bp mRNA EST 26-SEP-1997
LOCUS AA581944/c
DEFINITION nn40b07.s1 NCI_CGAP_GC5 Homo sapiens cDNA clone IMAGE:1086325 3',
mRNA sequence.
ACCESSION AA581944
VERSION AA581944.1 GI:2360622
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 431)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Insert Length: 1176 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 430.
location/Qualifiers
1..431
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1086325"
/clone_lib="NCI_CGAP_GC5"
/tissue_type="germ cell tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site.1: EcoRI; Site.2: XhoI
; Cloned unidirectionally. Primer: Oligo dT. Mixed germ
cell tumors. 5' adaptor sequence: 5' GAATTCGGACGACG 3' 3'
adaptor sequence: 5' CTCAGTTTTTTTTTTTTTTTTTTT 3' Average
insert size: 0.7 kb."
BASE COUNT 122 a 102 c 92 g 115 t
ORIGIN
Query Match 7.6%; Score 243.6; DB 10; Length 431;
Best Local Similarity 87.8%; Pred. No. 6e-44;

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 13:07:20 ; Search time 519.52 Seconds

(without alignments)
5290.623 Million cell updates/sec

Title: US-09-867-034-4

Perfect score: 3206

Sequence: 1 ttccgctcgcagctgaact.....ttgtatgagtgaaaaaaa 3206

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_1101:*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1030	32.1	2265	AA129139	Colon tumour relat
2	1007.4	31.4	2263	AA252937	Human prostate tum
3	655.6	20.4	1668	AA277886	Human cancer assoc
4	655.6	20.4	1668	AA277886	Human cancer assoc
5	654.4	20.4	1667	AA341197	Human colon cancer
6	365	11.4	536	AA341197	Human colon cancer
7	247.8	7.7	313	AA341197	Human colon cancer
8	200.2	6.2	417	AA341197	Human colon cancer
9	200.2	6.2	417	AA341197	Human colon cancer
10	187.6	5.9	401	AA341197	Human colon cancer
11	187.6	5.9	401	AA341197	Human colon cancer

12	148	4.6	457	22	AAH34198	Human colon cancer
13	136.8	4.3	295	16	AAH34198	Human colon cancer
14	123.2	3.8	593	21	AAH34198	Human colon cancer
15	116.8	3.6	620	21	AAH34198	Human colon cancer
16	109.4	3.4	570	21	AAH34198	Human colon cancer
17	105.8	3.3	936	22	AAH34198	Human colon cancer
18	105.8	3.3	936	22	AAH34198	Human colon cancer
19	105.8	3.3	936	22	AAH34198	Human colon cancer
20	105.8	3.3	936	22	AAH34198	Human colon cancer
21	105.8	3.3	936	22	AAH34198	Human colon cancer
22	105.8	3.3	936	22	AAH34198	Human colon cancer
23	103.6	3.2	936	22	AAH34198	Human colon cancer
24	103.6	3.2	936	22	AAH34198	Human colon cancer
25	103.6	3.2	936	22	AAH34198	Human colon cancer
26	103.6	3.2	936	22	AAH34198	Human colon cancer
27	103.6	3.2	936	22	AAH34198	Human colon cancer
28	103.6	3.2	936	22	AAH34198	Human colon cancer
29	53	1.7	5163	19	AAH34198	Human colon cancer
30	53	1.7	5163	19	AAH34198	Human colon cancer
31	53	1.7	5163	19	AAH34198	Human colon cancer
32	52.2	1.6	5318	19	AAH34198	Human colon cancer
33	50.4	1.6	5511	21	AAH34198	Human colon cancer
34	50.4	1.6	7334	21	AAH34198	Human colon cancer
35	48.4	1.5	867	20	AAH34198	Human colon cancer
36	48.4	1.5	867	20	AAH34198	Human colon cancer
37	48.4	1.5	1397	20	AAH34198	Human colon cancer
38	48.4	1.5	1397	20	AAH34198	Human colon cancer
39	47	1.5	244	22	AAH34198	Human colon cancer
40	45.6	1.4	5912	21	AAH34198	Human colon cancer
41	45	1.4	163	20	AAH34198	Human colon cancer
42	44.2	1.4	244	22	AAH34198	Human colon cancer
43	44.2	1.4	573	22	AAH34198	Human colon cancer
44	44.2	1.4	573	22	AAH34198	Human colon cancer
45	43.6	1.4	436	22	AAH34198	Human colon cancer

ALIGNMENTS

RESULT 1	AA129139	standard: cDNA: 2265 BP.
ID	AA129139	standard: cDNA: 2265 BP.
XX	AA129139	
DT	12-OCT-2001	(first entry)
XX	Colon tumour related longer determined cDNA sequence for C794P.	
DE	Human: immunotherapy; diagnosis; colon cancer; colon tumour;	
KW	immunogenic; gene therapy; vaccine; colonic cancer; ss.	
OS	Homo sapiens.	
XX	WO200149716-A2.	
PD	12-JUL-2001.	
XX	29-DEC-2000; 2000WO-US35596.	
XX	30-DEC-1999; 99US-0476296.	
PR	10-JAN-2000; 2000US-0480321.	
PR	15-FEB-2000; 2000US-0504629.	
PR	06-MAR-2000; 2000US-0519444.	
PR	19-MAY-2000; 2000US-0575251.	
PR	29-JUN-2000; 2000US-0609448.	
PR	28-AUG-2000; 2000US-0649811.	
XX	(CORI-) CORIXA CORP.	
XX	Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;	
PI	King GE, Wang T, Jiang Y;	
XX		

DR WPI; 2001-441847/47.
XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
XX
PS Claim 2; Page 324; 472pp; English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512
CC and AA24494 to AA24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
XX
SQ Sequence 2265 BP; 679 A; 484 C; 522 G; 580 T; 0 other;
Query Match 32.1%; Score 1030; DB 22; Length 2265;
Best Local Similarity 83.2%; Pred. No. 7e-276; Indels 145; Gaps 67;
Matches 2001; Conservative 0; Mismatches 260;
QY 801 ttgttcgtttttagaagggtattactacaactctctacatgtaagaaaggaaggtat 860
DB 2 ttgttcgtttttagaagggtattactacaactctctacatgtaagaaaggaaggtat 61
QY 861 tccctgagaagatttcagtgaacagatcagaacatttgaccagaagaagaacattcc 920
DB 62 tccctgtt-gaagatttcagtgaacagatcagaacatttgaccagaagaagaacattcc 120
QY 921 atggcctataaagactgcatagtaagtaattactagctgttttaagaatgtratttgaca 980
DB 121 atggcctataaagactgcatagtaagtaattactagctgttttaagaatgtratttgaca 180
QY 981 tctgtttatgagacactgtataatcttactgttaagcacactctgtccacaaagattctg 1040
DB 181 tctgtttatgagacactgtataatcttactgttaagcacactctgtccacaaagattctg 238
QY 1041 aaatggtgctgtatgagaagtttctgttaatgtaacaatgtaacaattttgacagaac 1100
DB 239 aaatggtgctgtatgagaagtttctgttaatgtaacaatgtaacaattttgacagaac 296
QY 1101 caaagatgaataatgagaagctgtgactgagaagaataataaagaactattataagtagc 1160
DB 297 caaagatgaataatgagaagctgtgactgagaagaataataaagaactattataagtagc 354
QY 1161 tcaagaacttttctaatacatgatgtgacactgtcggttgatgattgaagggcggg 1220
DB 355 tcaagaac-ttctaatacatgatgtgac-----ttcggtgtgattatattatgctg 407
QY 1221 aaccaagactgtgctgtagctgctcctcaatggttttagcactgcgaatgtgcaatgtcagc 1280
DB 408 taaccagactg--cgatgactgctcctcaat-ggttttagcgcga--ttgcaaatctgac 462
QY 1281 ctgcaagagcttaaccacagagcccttctgtgcttccagcttccaaagtctctga 1340
|||||

DB 463 ctgcaagagcttaaccacagagcccttctgtgcttccagctca-agtgtcctga 521
QY 1341 tgcctgaacgcagacagacagaacgaatgcttaataagaagatggtgggtccctgca 1400
DB 522 tgcctgaacgcagacagacagaacg-aatgcttaataagaagatggtgggtccctgag 580
QY 1401 gttgtgctgtggtcccggtctacacagaagatgtaatgtgggaactgtcacaagtgtgc 1460
DB 581 gttgtgctgtggtcccggtctacacagaagatgtaatgtgggaactgtcacaagtgtgc 635
QY 1461 attggtctcaagtgagctgactgactgtagagacaatttcgctgacttactatttgc 1520
DB 636 attt-ggtctcaagtgagctgactgactgtagagacaatttcgctgacttact--attgc 692
QY 1521 gggcaccatgctgtgcatcttcatctccagatgataattgcatgtgtctcactaa 1580
DB 693 gggcaccatgctgtgcatcttcatctccagatgataattgcatgtgtctcactaa-agcaa 751
QY 1581 gatcaatacaaaaagcgaagcatattgaaagaacggaactgtattgaagaactttca 1640
DB 752 gatcaatacaaaa-cgaagcatattgaaaga-gagaactgtattgaagaactttca 809
QY 1641 aaatctaaactggtgcgacagagcttcacacatctatgagcatagcgaagcttctc 1700
DB 810 aaatctaaactggtgcg-acaggtctcacacatct-tgagacagaagagagcgtcttc 867
QY 1701 cctcaagtcagattagagctcctcacaagaagcgcctagatgacaaaatcccgtagttca 1760
DB 868 cctcaagtcagattagagctcctcacaagaagcgcctagatgacaaaatcccgtagttca 921
QY 1761 agacacagcagatgccccggcctgactatagaaatcagaaatgtgaaacccgca 1820
DB 922 agacacagcagatg-cccggcctgactatagaaatcagaaatgtgaaacccgca 979
QY 1821 tggcccccaacatattgataatattatctcagagatgtttagaagactgtcag 1880
DB 980 tggcccccaacatattgataatattatctcagagatgtttagaagactgtcag 1033
QY 1881 agaagtagacacacgtaaaagactgtgctccggagatgtctctctccatcagacata 1940
DB 1034 agaagtagacacac-gtaaaagatcgtgctccggagatgtctctctccatcagacatgc 1091
QY 1941 ctgcagctcctcatctgcaatggaacggtgtgcaatgtctgcacaacgacatccagct 2000
DB 1092 cagctctctcgaatggaagttgtgaatgttgcga-----cgaaatccagct 1137
QY 2001 cactgtgtaaaatgaagatctatgacatttaacatgtagctcgatgctattagcgctgc 2060
DB 1138 cactgtct-aaataagaactatgacatca--aatgtagtagatgctattagcgctgtc 1194
QY 2061 tcagagaggtgtggttttctcctaactcagtaacaaagtaactgagacaatgcttaagggttgc 2120
DB 1195 agaga--ggtgtgtttcttcaatcagt-acaaagtagtagacaatgcttaagggttgc 1249
QY 2121 ttctctaattctttccctgtgtgaggaacaagaacccatttccaatctcagaaggaagc 2180
DB 1250 ttctctaattctttt-cctgtgtgaggaacaagaac--atttccaatctcagaaggaag- 1305
QY 2181 ctccccagactgctgtgctgccttgaggccaacacatgctcttgagttgttgaactaa 2240
DB 1306 ctccccagactgct-----tgcttcgtggaacaacatgctcttgagttgaag-tgaacctaa 1360
QY 2241 ctccccctggagacacatacagcatcaacggtgaggttcggaagggtatgagaagagat 2300
DB 1361 ttccccctggagacata---cgatacaactgtgaggttcggaagggt--atgagaagagat 1415
QY 2301 acccaacatcttcaataggttcaagaactacactctcgttgaacaatcagaatcagaatgga 2360
DB 1416 acccaacatcttcaaggttcaagaactacactctc--tgacaatcagaataggg--ac 1470
QY 2361 cgtgtcttatacctccaagaagagatctgtgccaaccccccttttttgaataaaca 2420
DB 1471 accgtcttatacctccaagaatga-gagatctgtgca-----ccttga 1513
|||||

OY	2421	gcccccaagagcttggaaacctgagcctcaacccaagaagactggaaagaaacatct	2480
Db	1514	acagccagaagcctt-gaaaccttagcctc--acccaagaagaaactggaaa--gaagacat	1569
OY	2481	ttcagctcttttcaggagcgctgctctgggaatccagaaacgctttttga tgcataatga	2540
Db	1570	ctaagctcttttcaggagcgctgc--tgggaatccaggaaac--tlttgatgtctaattga	1627
OY	2541	ggcctggaactataataatgctcactctatcgtgggtttaaactcaagltttgaacatgcta	2600
Db	1628	ggcctggaactataaataatgctcactc---tatgggggtgcactctcaaglttttgaa--atgcta	1682
OY	2601	ggagagcgaaacgggggcccagaggaataaaaaaataatgacctggtagaagaagaagggc	2660
Db	1683	ggagagcgaaagggggcagagag---taaaaaaataatgacctggtagaagaagaagggc	1739
OY	2661	ggaacacgggttggggaagatcaatltgaagaaggaggaaccttggatccaacttgcctt	2720
Db	1740	ggaacacgggttggggaagatcaatltgaagaaggaggaaccttggatccaacttgcctt	1798
OY	2721	aggctccctccctccatgacagcaaaaggacatctctctaaatcattgcctccccaagactg	2780
Db	1799	aggctccctccctccatc--caacaaaggagcaactctctaa--tcaatgcctccccaagactg	1856
OY	2781	gcttggggaagaggttttaaaaaaacaataatccagggatgaaggagccttgggttcagtttg	2840
Db	1857	gcttggggaagaggtt--aaaaaacaataatccaggaagt--aagagccttagg---tcaagtt	1911
OY	2841	aaaattggaaacaacaacttgcctctggcgaagaagggtbccaaagacggaagcttgttgc	2900
Db	1912	gaaattggagagacaact--gtctggcgaagaagggtgc---gaaggggagacttgttgc	1966
OY	2901	gtcccaagcgttccagcctcggggtgtgaaggtctctgaaggtgtgtccatvggggagc	2960
Db	1967	gt--ccaagcgttccagcctcggggtgt--aggttctgaaggtgtgtccatvggggagc	2024
OY	2961	ttctctgttaccggagagcctcaacttgtgcacaaacaaacaaacaaacaaacaaacaa	3020
Db	2025	ttctctgttaccga--gaaggtccagctgtgtgcacaaacaaacaaacaaacaaacaa	2083
OY	3021	cacacacaagaatggggggaaccccaatccacgtacccagttacaagaacttacaacaagt	3080
Db	2084	cacacacaagaatggggggaaccccaatccacgtacccagttacaagaacttacaacaagt	2137
OY	3081	tgtcccttttaatttctaataagcctgtccctcttaaagtaatttattgtattat	3140
Db	2138	tgtcccttttattatccaat--gcctgtccctcttaaagtaatttattgtattat	2196
OY	3141	tttgtcttaccgttcaattgttaattgtaattgaatgaataaaggccttggtagatgtgaa	3200
Db	2197	tttgtcttaccgttcaattgttaattgtaattgaatgaataaaggccttggtagataaaaaa	2256
OY	3201	aaaaaa 3206	
Db	2257	aaaaaa 2262	
RESULT 2			
AAZ52937 standard; cDNA; 2263 BP.			
XX	AAZ52937:		
XX	AAZ52937:		
XX	AC		
XX	14-MAR-2000 (first entry)		
XX	DT		
DE	Human prostate tumor cDNA library derived EST fragment #80.		
XX	Pancreas; tumor; EST; expressed sequence tag; human; cytosolic;		
KW	treatment; ds.		
XX	XX		
XX	Homo sapiens.		
XX	OS		

PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR MPI; 1999-621386/54.
XX
P-PSDB; AAY74052, AAY74053, AAY74054.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
XX proteins -
PS
XX
XX Claim 2; Page 247-248; 502pp; German.
XX
XX This invention describes novel polypeptides and their encoding nucleic
XX acids derived from human pancreatic tumor tissue which have cytostatic
XX activity. The sequences are also useful in producing pharmaceutical
XX compositions for treatment of pancreatic tumors. AAZ52858-253014
XX represent expressed sequence tag (EST) fragments derived from a human
XX pancreatic tumor cDNA library and which encode the proteins represented
XX in AAY3814-VY4252.
XX
XX Sequence 2263 BP; 689 A; 484 C; 520 G; 570 T; 0 other;

Query Match	31.4%;	Score 1007.4;	DB 20;	Length 2263;
Best Local Similarity	83.0%;	Pred. No. 1.3e-269;		
Matches 1979;	Conservative	0;	Mismatches 261;	Indels 145; Gaps 67;

QY	882	atctctcaactctctcaactgtgaagaagaaggaagtattccctcgtgagaagattctcaatg	881
Db	1	attagaanaactctctcaactgtgaagaagaaggaagtattccctcg-gaagattctcaatg	59
QY	882	acagatatacagaacatttgcgccagagaagaacacatctcagtcgcctatacaacttgat	941
Db	60	acagatatacagaacatttgcgccagagaagaacacatctcagtcgcctatacaacttgat	119
QY	942	agtgaatctactagctctgtttaaagatgfatlttggcaactctgttataygacagactgta	1001
Db	120	agtgaatctactagctctgtttaaagatgfatlttggcaactctgttataygacagactgta	179
QY	1002	attcttacttgaagcgcacatctcgtacacaagaattctgaattgcgtgtttatgaacag	1065
Db	180	attcttacttgcac-gcacacctcgttccacaaga-tctgaattgcgtgc-tyatgacaaag	236
QY	1062	tttgtttaatgtaacaaataglaaacatttttggagagaacaccaaagtgaacaatgagaagac	1121
Db	237	-tttgtttaatgtaacataagtaaacatttttggagagaacaccaaagtgaacaatgagaagac	295
QY	1122	tgtgactgagagaataataaagaacattatataagtagtactgaagcaacttctctaacta	1181
Db	296	tgtgact-gagagaataataataaagaac-atagaagtagtactgaagcaac-ttctctaacta	352
QY	1182	tgatgagacccctgcgtgtgtgattgtatgaagcctggagaaacaaagctgctgtatgac	1241
Db	353	tgatttggcccttc-----ggtgtgatatctatgtgcgttaacagacgcg--ggatgac	404
QY	1242	tgcctcaatggtgttagcagatgcatgtgcaaatgctgacccctgcaaaagcctaacccacag	1301
Db	405	tgcctcaat-ggtttagcagatgcatgttc--aaatcgtgacctgcaaaagcctaacccacag	461
QY	1302	agcccttctgcgctgtgttcacgtccacagatgtcctctgtgtcctgcaagcagacagaca	1361
Db	462	agcccttctgcgctgtgttcacagctcaca-agtgcctgcatgctcgtgaacagcagacaca	520
QY	1362	gagaatgtctlaataaagaagatgtgtgtgttccctcgtcagtttgcgtgtgcgtccggt	1421

Db	512	-----cgaatccagctcacttgcgt--aaataagaactatgatgacatla--aatgtatgt	559
Qy	2041	cgatgctattatagcgctgtgcctcagagagtggtgtttcttcacatcagtlacaaagttacgt	2100
Db	560	agatgctattatagcgctgttcagaga---ggttgttttcttcacatcagtt--acaaagtactgt	615
Qy	2101	agacaatgcttaagggtgtgtttcttcataatctttccctgtgtatggccaagaccocat	2160
Db	616	agacaatgctta--gggtgtgtttcttcataatctttt--cctgtgtatggccaacagaacc--a	671
Qy	2161	ttccaatcctagagaagaagcctcccccagcatgtcctgttcctcctgtggccaacactgtct	2220
Db	672	ttccaactctagagaagaag--ctcccacagatgcgt---tgcctcctgtggccaacatgtcgt	726
Qy	2221	cttgatgttaagtcttaccatacttccccgtggagacataccgcatcactaactgtgtgtgtctc	2280
Db	727	cttgatgtttaag--tgacttaattccccctgtggagagacata---cgcatcactatgtgtgtgtc	782
Qy	2281	gaggtggatgtgagaagaaggataccacacatcttcataatgtgttcaagaagtacactctgtg	2340
Db	783	gaggtgg--atgagagaaggatataccacacacttctcaaggtgtcaacaagtctactctc---tg	837
Qy	2341	acaagtccagatlaaggagacaacactgtctctatctcccttcacatgtgagagatctgtggccaac	2400
Db	838	acaagtccagatlaagg--acagtgtctctatctcccttcacatgtga--gagattctgtgaca---	891
Qy	2401	cccccttttttggaaaaaccggccccccagagctctgtgcaacctttagccttcacacagaaga	2460
Db	892	-----ccttggacaagccccagagctct--gcaaccttagcctc--accacaagaaga	936
Qy	2461	cttgaaaaaggagacatalactttcttcagctctttctcagagaggtgtcctgtggatccagagac	2520
Db	937	cttgaaaa--gagacatatc--tctcagctctttctcagagaggtgtcct--tggatccagagac	993
Qy	2521	gtttttgtatgtctaattcagaagggcctgtgactataataatgtccaactatavtgggttttaac	2580
Db	994	tttttgatgtctaattcagaagggcctgtgactataataatgtccaactat---gggtgtcactc	1048
Qy	2581	tacagtttttgaataatgtctagagagcgagagcgagagcgagaggttaaaaaaatatgacctgt	2644
Db	1049	tacagtttttgaaa--atgtcctagagagcgagagagggcgagagag---taaaaaaatatgacctgt	1104
Qy	2641	gtagaagaaggagaagagcgaaaagaactgtgtgtgtgtgtgtatcattagagagagggcacct	2700
Db	1105	gtagaagaaggagaagagcgaaaagaactgtgtgtgtgtgtgtatcattagagagagggcacct	1164
Qy	2701	gggatccacacttgcgttctcctaagtctccctctccatgtcagcaaaaggacattctcctaag	2766
Db	1165	gggatccacacttct--tctcctaagtctccctctccatc--cagcaaaaggagacattctccta--	1221
Qy	2761	tcatgtccctcccggaagcctgtcgtgtgtgtgtgtgtttaaanaacaaaatccagaggttaa	2822
Db	1222	tcatgtccctcccggaagcctgtcgtgtgtgtgtgtgtgt--aaaaacaaaatccagaggt--aa	1279
Qy	2821	gagccttaagggttccagtttttgaanaatltvgagagacaacatgttctgtgcgaagaaggtgtgcaga	2880
Db	1280	gagccttaagg---tccagtttgaanaatltvgagagacaactgtc--ctvgcaaaaggtgtgc--ga	1331
Qy	2881	gagagagcttgttgcctcagagatccccagcgttccagcctcaggggtgttaagtctctgaagt	2940
Db	1332	gagagagagcttgttgcctcagagat--ccagcgttccagcctcaggggtgtc--aggttctcgaagt	1389
Qy	2941	gtgcacatvgggggtccctcagcctctcgtgtgaccggaggtctcagcctgtgtggccacaaca	3000
Db	1380	gtgcacatvgggggtccctcagcctctcctcgtgtga--cagaggtctcagcctgtgtggccacaaca	1448
Qy	3001	caaccacacacacacaacacacacacacaaaatgtggggccaaccacatccacgttaaccaagc	3060
Db	1449	caaccacacacacacacacacacacacaaaatgtggggccaaccacatccagttacaagc---	1504
Qy	3061	tttaacacaaatgttatagtgtcccttttattcttcaatagcctgttccctcttaaaagt	3120

Query Match	20.4%	Score 654.4	DB 22	Length 1667
Best Local Similarity	81.0%	Pred. No. 1.6e-171		
Matches 1431: Conservative	2	Mismatches 213	Indels 120	Gaps 53

OY	1441	gggaactgcgcaaaagtgtgcatttgggtctcacgttggaactcgagctctaaagaaataattca	1500
Db	1	gggaactgcgcaaaagtgtgcattt-gggtctcacgttggaactcgagctctaaagaaataattca	59
OY	1501	gctgctccatactatttgggtgggacacatcgctgcgcatctgtcatctccaaagtataat	1560
Db	60	gctgctccatact--attgtyggacacatctgcgcgcatctgtcatctccaaagtataat	117
OY	1561	gcatgtattgtcactatgacaagatccaataaacaagcgaagcatatgtaagaacagagac	1620
Db	118	gcatgtattgtcac-agaagatccaataaacaana--cgaagcatattgaaga-gagagac	174
OY	1621	tgtattgacgaagagcttccaataatctcaaatcgaatcgtgtcgaagcttccacaactctg	1680
Db	175	tgtattgacgaagagcttccaataatctcaaatcgtgtctg-acagagcttccaactct-tg	232
OY	1681	gagcatacgaagagcttctccctcgaagtcagagatttcacgctccgaagagacgcgttaat	1740
Db	233	gagcatacgaagagagcttcttctccaaagttcagagtaacggtccca--gagacagcgaag	290
OY	1741	gcaaaaatcccgtagtttcaagaacagacagcatgccccggcgctgactatgaatcca	1800
Db	291	gcaaaaatccc---tatccaagmcaagcagcatg-ccccgccttgacattatgat-ca	344
OY	1801	tcagaatgttggaacccgcgcctatggcccccaacataatgatcatatctattctcagcgt	1860
Db	345	taagaatgttggaacccgcgcctatggcccccaacaaatgtacaaactatattt---aga	398
OY	1861	gtttagacaagaacatcgcatctggagagatgagacacagtaaaagactctggtcccggaatt	1920
Db	399	gtgtttagaagaagactgtgttggaagatgagaccca--gtaagaatctgmcctcggtttt	456
OY	1921	tctcttccatctagacatactctgccagctccatactctgcaatgagcaagcttgtgcaatgct	1980
Db	457	tctctt--catctgacatctgcgcagcctctctgtgaatggaagttgtgaattgtgcaag--	512
OY	1981	tgtgcaaacgacatccacgcgcactctgtctaataaagaactctatgacattaaatgtagct	2040
Db	513	-----aatccagctccacttgcct--aataagaactctatgacatta--aatgtgct	558
OY	2041	cgatgctattagcgtctgtgctcagagaggttggttttcttcataatcagtaacaagctagct	2100
Db	559	agatgctattagcgtctgttcagaga--ggtgtgttcttcataatcagt-acaaagtactg	614
OY	2101	agacaatgcttaggggtgtgtttcttcaattcttctccctggtgaaggcaaaagcccat	2160
Db	615	agacaatgcttta-gggttgtttcttcaattcttct--cctgtgaaggcaaaagacact	672
OY	2161	tctcaaatctagaagaagcctccccaagcatctgtctgtctccctgggccaacactgtt	2220
Db	673	t-----ccaatccagaggaag-ctcccagcatgtct---tgtctctgggcaaacattgtct	725
OY	2221	cttgagttaagtgtgacctaaactccctctgggaagacatacgcgcatcaactgttggagctcc	2280
Db	726	cttgagttaag-tgacctaatctccctctgggagacata--cgatccaactgttggagctcc	781
OY	2281	gagggtgatatgaaaggtataccaccacactcttcataggttcaacagcttaacactctgtg	2340
Db	782	gagggtg--atgaaagaaaggtataccaccacactcttcaaggggttcaacagcttaacctc--tg	836
OY	2341	acaagtcagaatagggaacaccttgccttcaatccctccaattgaaagagcttctggccaac	2400
Db	837	acaagtcagaatagggt--acaatgcttcatctccctccaattgaa-gagattcttggaaa---	890
OY	2401	cccccttlttttgaaacacgagcccccagagctgttgcaacctagcctcaaccagaaga	2460
Db	891	-----octtttgaaacagcccaagactt-gcaacctagcttc-accacaaga	935
OY	2461	cttgaaagagagacatacttctaagctttttcaagggcggtgctcttgggaatccaaggaac	2520
Db	936	cttgaaac-gagacatactc-ctctaagctttttcaagggcggttgcct-ttgggaatccaaggaac	992

[illegible]

PI Baker KP, Goddard A, Wood WI;
XX
XX WPI: 2001-112729/12.

PT New isolated nucleic acid molecule encoding a SRT polypeptide is useful
PT for production of recombinant SRT polypeptides, gene mapping,
PT diagnosing genetic disorders and for gene therapy -

PS Claim 2; Fig 401; 663pp; English.

CC Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding
CC human SRT proteins. The cDNA sequences are isolated from various
CC different human tissue cDNA libraries. The invention relates to a method
CC for detecting cDNA encoding an SRT protein, a vector containing cDNA
CC encoding SRT, a host cell transformed with the vector, an isolated SRT
CC polypeptide, and an antibody which binds to SRT. The polynucleotide
CC sequence can be used in gene therapy and is useful in the recombinant
CC production of SRT polypeptides, as a hybridisation probe to screen
CC libraries to isolate cDNAs with sequence identity to SRT polypeptides, to
CC map the gene encoding the SRT polypeptides and analysing genetic
CC disorders, tissue typing and disease tissue detection. The SRT
CC polynucleotide sequences can be used in polymerase chain reaction,
CC screening for new therapeutic molecules and generation of antisense RNA
CC and DNA.

SO Sequence 536 BP; 166 A; 156 C; 77 G; 134 T; 3 other;

Query Match 11.4%; Score 365; DB 22; Length 536;
Best Local Similarity 90.5%; Pred. No. 2.6e-91;
Matches 503; Conservative 0; Mismatches 33; Indels 20; Gaps 10;

QY 279 taacacacacagaagaactgactggtctctacagtcgacgctgataccactgaaa 338
DB 1 taacacacacagaagaactg-gantagtggtctctacagtcgacgctgataccactgaaa 59
QY 339 ctaattgacctgaactgctagcaccacagaacaataacactctctctccacagctact 398
DB 60 ctaattt-ccttgaactgctagcaccacagaacaataacactctctctccacagctact 118
QY 399 tcacctgtcccccataataatagtaacataagttctctccacaattcctaccgtctcc 458
DB 119 tcacctgtcccccataataatag-acacatagttctctccacaattcctaccgtctcc 177
QY 459 cccataatagtaacataatgctccctccacaattcctaccgtctgacgacagtgag 518
DB 178 cccataatagtaacataatgctccctccacaattcctaccgtctgacgacagtgag 237
QY 519 tcaaccacaatgtaaatcagttagctactctgacataatcacgcgttcaatccaaa 578
DB 238 tcaaccacaatgtaaatcag-ttagctactctgacataatcacgcgttcaatccaaa 296
QY 579 tgatggtataatcaatagttctctctgnaacacaagtaacataatgaatgtcccca 638
DB 297 tgatggtataa-tcaacatggtctctctgnaacacaagtaacataatgaatgtcccca 355
QY 639 ccacagaagaacaatcaatccctgagtcctccactggacccggttattattgagtgacc 698
DB 356 ccacagaagaacaatcaatcalcag-ggctccactgacgacgcttatttggagccagc 414
QY 699 atgcaagcctaaacagcagtgctccagcaatccttgccaagaatgatacccccgtgca 758
DB 415 a-----ccctaacaagcagag-gtccacagcaatccttgcca---agatgatacccccgtgca 465
QY 759 gataattggtatggttggtaagcttgcataataacagttttgctgtgttaagaag 818
DB 466 gataattggtatggttggtaagctgc-----ataatacaagttttgctgtgttaagaag 520
QY 819 ggtattactaacactc 834
DB 521 ggtattactaacactc 536

RESULT 7
AAV87786
ID AAV87786 standard; cDNA: 313 BP.

XX
XX AAV87786;

DT 12-FEB-1999 (first entry)

DE EST clone ES206.

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

PN W09845437-A2.

PD 15-OCT-1998.

PF 10-APR-1998; 98WO-US06956.

PR 10-APR-1997; 97US-0837312.

PA (GEMV) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PT Racine LA, Spaulding V, Treacy M;

DR WPI: 1999-070078/06.

PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries

PS Claim 1; Page 179; 641pp; English.

CC The present sequence represents an expressed sequence tag (EST), and is
CC a polynucleotide of the invention. The polynucleotides of the invention
CC are all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.

SO Sequence 313 BP; 95 A; 103 C; 43 G; 72 T; 0 other;

Query Match 7.7%; Score 247; 8; DB 20; Length 313;
Best Local Similarity 98.5%; Pred. No. 7.4e-59;
Matches 271; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 192 taaatgaagaagcattatcatctactctctgtgctctctctctctgtaaacacagcca 251
DB 37 taaatgaagaagcattatcatctactctctgtg-ctctctctctgtaaacacagcca 95
QY 252 ccaaccagaagcaactcagctgattgctgtaacacacacagaagaactgagctagtgtccta 311
DB 96 ccaaccagaagcaactcagctgattgctgtaacacacacagaagaactgagctagtgtccta 155
QY 312 cagtagctgcagctgataccactgaactaatgtgcccctgaaactgctagacacacagca 371
DB 156 cagtagctgcagctgataccactgaactaatltt-ccttgaagaactgctagacacacagca 214
QY 372 aataacacttcttccacacagctacttcaactgtctcccccataatagtaacacatagt 431

XX PS Claim 2; Page 168; 472pp; English.

CC XX The present invention describes colon tumour associated proteins (I) and

CC the polynucleotides (II) that encode them. (I) have cytostatic activity.

CC (I) and (II) can be used in gene therapy and vaccine production. (I) and

CC (II) may be used in the prevention, diagnosis and treatment of diseases

CC associated with inappropriate colon tumour associated protein (TCAP)

CC expression, such as colonic cancer. For example, (I) and (II) may be

CC used to treat disorders associated with decreased expression by

CC rectifying mutations or deletions in a patient's genome that affect the

CC activity of TCAPs by expressing inactive proteins or to supplement the

CC patients own production of them. Additionally, (II) may be used to

CC produce the TCAP proteins, by inserting the nucleic acids into a host

CC cell culturing the cell to express the protein. (II) and its

CC complementary sequences may also be used as DNA probes in diagnostic

CC polymerase chain reaction (PCR) and hybridisation assays to detect and

CC quantitate the presence of similar nucleic acids in samples, and

CC therefore which patients may be in need of restorative therapy. (I) may

CC also be used as antigens in the production of antibodies against TCAPs

CC and in assays to identify modulators of TCAP expression and activity.

CC Anti-(I) antibodies and antagonists may also be used to down regulate

CC TCAP expression and activity. The anti-(I) antibodies may also be used

CC as diagnostic agents for detecting the presence of TCAPs in samples

CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512

CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences

CC given in the exemplification of the present invention.

XX SQ Sequence 417 BP; 87 A; 126 C; 88 G; 116 T; 0 other;

Query Match 6.2%; Score 200.2; DB 22; Length 417;

Best Local Similarity 88.3%; Pred. No. 1.5e-45;

Matches 356; Conservative 0; Mismatches 28; Indels 19; Gaps 12;

QY 2570 gggtttaattacagtttttgaaacatgtagagagcagagccagagagtaaaaa 2629

DB 394 GGGGTGACACTCTACAGTTTGGAA-ATGCTAGAGAGGAGGAGGACAGAG--TAAAA 339

QY 2630 aacatgacctgtagaagaagaagaagaagaacagggtgggagagatcaatagaag 2689

DB 338 AACATGACCTGTGTAAGAAAGAGAGAGGCAAAAGAAATAGTGGGAGATCAATTAGAG 279

QY 2690 agaggagacccggagatcaccttcgttccttaggtccctccatgcagcaaaagagac 2749

DB 278 AGGAGGACCTGGGATTCACCTTC-TTCCTTAGAGTCCCTCTCTCAT-CAGCAAAAGAGC 221

QY 2750 actctcttaagtcacgtccctccgaagacgtgtagaagaagtttaaaaaaacaanaat 2809

DB 220 ACTTCTCTTAA-TCAATGCCCTCCCGAAGACTGCTGGAGAAAGTTT-AAAAAAANAAT 163

QY 2810 ccaagagaagaagccttagagtcagtttgaataatggaacaaactgcttgagaaa 2869

DB 162 CCAGAGAGT-AAAGACCTTAGG---TCACTTTGAAATTTGCAACAAACT--GTCTGGCAAA 109

QY 2870 ggggtcccaagagcgagctgtgtcctcagagatcccaagcgtccagcctcgaggtgtaag 2929

DB 108 GGGTGC---GAGAGGAGACTGTGCTCAGAGT--CCAGCCTCCAGCCTCGGGGTGT-AG 54

QY 2930 gctctgaggtgtagcagatggagcctcagccttctgtgtac 2972

DB 53 GTTCTGAGGTGTGCATTTGGGCTCAGCCTTCCTCTGTGTGAC 11

RESULT 10

AAA77762

ID AAA77762 standard; cDNA; 401 BP.

XX AC AAA77762;

XX 14-NOV-2000 (first entry)

DE cDNA encoding human colon tumour polypeptide, SEQ ID NO:41.

XX KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine;

KW immunotherapy; diagnosis; progression; ss.

XX OS Homo sapiens.

XX PN WO200037643-A2.

XX PD 29-JUN-2000.

XX PF 23-DEC-1999; 99WO-US30909.

XX PR 23-DEC-1998; 98US-0221298.

PR 02-JUL-1999; 99US-0347496.

PR 22-SEP-1999; 99US-0441064.

PR 19-NOV-1999; 99US-0444242.

PR 02-DEC-1999; 99US-0454150.

XX PA (CORI-) CORIXA CORP.

XX PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk J;

PI Wang T, Yugu J;

XX DR WPI: 2000-442671/38.

XX PT New colon tumor polypeptides used to inhibit the development of cancer,

PT especially colon cancer, and for diagnosing and monitoring the

PT progression of the cancer -

XX PS Claim 1; Page 99; 229pp; English.

XX CC Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or

CC portions of proteins which are associated with human colon tumours.

CC The invention also specifically discloses 8 human colon tumour proteins

CC (AAB11897-B11904). The nucleic acids, the polypeptides they encode, and

CC antigen presenting cells (APCs, preferably dendritic cells) expressing

CC such polypeptides may be used in vaccines that target tumour cells,

CC especially colon tumour cells, thereby inhibiting the development of

CC cancer. T-cells specific for the polypeptide expressed by the APC are

CC used to remove tumour cells from biological samples, especially blood or

CC fractions thereof. The sample or the isolated T-cells specific for the

CC polypeptide can then be used to inhibit cancer development. CD4+ and/or

CC CD8+ T-cells from a patient may be incubated with a polypeptide or

CC nucleic acid of the invention, or an APC expressing such a polypeptide,

CC to cause the proliferation of specific T-cells. The T-cells can be

CC cloned and then administered back to the patient to inhibit cancer

CC development. Nucleic acids encoding the polypeptides and antibodies

CC against the polypeptides may be used to determine the expression level

CC of a tumour protein of the invention, and therefore to determine whether

CC cancer cells are present. Such diagnostic methods may also be used to

CC monitor the progression of a cancer by repeating the processes at time

CC intervals, and comparing the current result to previous results. The

CC present sequence represents a cDNA encoding a human colon tumour

CC polypeptide.

XX SQ Sequence 401 BP; 113 A; 85 C; 114 G; 85 T; 4 other;

Query Match 5.9%; Score 187.6; DB 21; Length 401;

Best Local Similarity 85.7%; Pred. No. 4.6e-42;

Matches 336; Conservative 0; Mismatches 38; Indels 18; Gaps 11;

QY 2570 gggtttaattacagtttttgaaacatgctgtaggagcagagggccaagagataaaaa 2629

DB 23 ggggtgacactcaagtttttgaa-atgctagagagcagaaggcagagag---taaaa 78

QY 2630 aacatgacctgtagaagaagaagaagaagaagaactgggtgggagagatcaattagaag 2689

DB 79 aacatgacctgtagaagaagaagaagaagaagaagaactagtgtaggagatcaattagaag 138

QY 2690 agaggagcaactggagatccacttcctcttagtgatccctccatcagcagcaagaagagc 2749

DB 139 agaggagcaactggagatccacttcctcttagtgatccctccatcagcagcaagaagagc 196

QY 2750 actctctaaagtcaccccccgaagactgctgagagaaggtttaaaaaaacaanaat 2809
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 197 actctctaa-tcatgcccctccgaagactgctgagagaagttt-aaaaaacaanaat 254
 QY 2810 ccaggagtaaaagccttagtgcagtttgaataatggagacaactctcttgcaaa 2869
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 255 ccaggagtt-aagagccttaag---tcagttgaattgagacaact--gtctggcaaa 308
 QY 2870 ggggtccaaagcggagctgttgcctcaaggtccagagctccgaggtgtaag 2929
 ||||||| || || ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 309 ggggtgc---ganaggagagctgtgctcangagtcacagcccgtccagcctggggtgt-an 364
 QY 2930 gtctctgaaggtgtgcacatggggagcctcagcct 2961
 || ||||| ||||||| ||||||| || ||||| |||||
 Db 365 gtctctgaaggtgtgcacatggggagcctcagcct 396
 RESULT 11
 AAI28500
 ID AAI28500 standard; cDNA: 401 BP.
 XX
 AC AAI28500;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Colon tumour related cDNA sequence Contig 42.
 XX
 KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 XX immunogenic; gene therapy; vaccine; colonic cancer; ss.
 OS Homo sapiens.
 XX
 PN WO200149716-A2.
 XX
 PD 12-JUL-2001.
 XX
 PE 29-DEC-2000; 2000MO-US35596.
 XX
 PR 30-DEC-1999; 99US-0476296.
 XX 10-JAN-2000; 2000US-0480321.
 PR 15-FEB-2000; 2000US-0504629.
 PR 06-MAR-2000; 2000US-0519444.
 PR 19-MAY-2000; 2000US-0573251.
 PR 29-JUN-2000; 2000US-0609448.
 PR 28-AUG-2000; 2000US-0649811.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Lodes MJ, Secret H, Benson DR, Meagher MJ, Stolk JA;
 PI King GE, Wang T, Jiang Y;
 XX
 DR WPI: 2001-441847/47.
 XX
 PT Colon tumor associated proteins and nucleic acids useful for the
 XX prevention, diagnosis and treatment of colonic cancer -
 XX
 PS Claim 2: Page 141; 472pp; English.
 XX
 CC The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytosolic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patients own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and

CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify mediators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
 CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
 CC given in the exemplification of the present invention.
 XX
 SQ Sequence 401 BP; 113 A; 85 C; 114 G; 85 T; 4 other;
 Query Match 5.9%; Score 187.6; DB 22; Length 401;
 Best Local Similarity 85.7%; Pred. No. 4.6e-42;
 Matches 336; Conservative 0; Mismatches 38; Indels 18; Gaps 11;
 QY 2570 ggggtttaactacagtttttgaacatgtctagaagcagaagcgagcagataaaa 2829
 ||| | ||||||| ||||||| ||||||| || ||||| |||||
 Db 23 ggggtgcactctacagtttttgaa-atgctagagagcagaagggcagagag---taaaa 78
 QY 2630 aacatgaacctgtgagagagagagagcagaagaaactgggtggagagatcaattagag 2889
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 79 aacatgaacctgtgagagagagagagcagaagaaactaggtggagagatcaattagag 138
 QY 2690 agagagcaccctggatccacacttgccttagtccctccctcatalgcagcaaaagagc 2749
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 139 agagagcaccctggatccacacttcccttagtccctccctccat-cagcaaaagagc 196
 QY 2750 actctctaaagtcaccccccgaagactgctgagagaaggtttaaaaaaacaanaat 2809
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 197 actctctaa-tcatgcccctccgaagactgctgagagaagttt-aaaaaacaanaat 254
 QY 2810 ccaggagtaaaagccttagtgcagtttgaataatggagacaactctcttgcaaa 2869
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 255 ccaggagtt-aagagccttaag---tcagttgaattgagacaact--gtctggcaaa 308
 QY 2870 ggggtccaaagcggagctgttgcctcaaggtccagagctccgaggtgtaag 2929
 ||||||| || || ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 309 ggggtgc---ganaggagagctgtgctcangagtcacagcccgtccagcctggggtgt-an 364
 QY 2930 gtctctgaaggtgtgcacatggggagcctcagcct 2961
 || ||||| ||||||| ||||||| || ||||| |||||
 Db 365 gtctctgaaggtgtgcacatggggagcctcagcct 396
 RESULT 12
 AAH34198
 ID AAH34198 standard; cDNA: 457 BP.
 XX
 AC AAH34198;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:1280.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 XX colorectal carcinoma; ss.
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PE 28-SEP-2000; 2000MO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX

DF 14-JUN-2000 (first entry)
 XX Human colon cancer differentially expressed nucleotide sequence #132.
 DE
 XX
 KM Colon cancer: detect; differential expression; human; treatment;
 KM detect mutation; non-invasive diagnostic method; ds.
 XX
 OS Homo sapiens.
 PN WC200012702-A2.
 PD 09-MAR-2000.
 XX
 PF 30-AUG-1999; 99WO-US19424.
 XX
 PR 31-AUG-1998; 98US-0098639.
 PR 27-JAN-1999; 99US-0117393.
 XX
 PA (FARB) BAYER CORP.
 XX
 PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
 PI Catino TJ, Dwiwedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
 PI Schlegel R;
 XX
 DR WPI; 2000-256641/22.
 XX
 PT Novel nucleic acids and proteins for identifying therapeutic agents
 PT useful for treating and diagnosing cancer, especially colon cancer
 XX
 PS Claim 16; Page 177; 345pp; English.
 XX
 CC This sequence represents a human nucleotide sequence which is
 CC differentially expressed in colon cancer cells compared to the expression
 CC levels in normal cells. The nucleotide sequence can be used as a source
 CC of primers and probes. The nucleotide sequence is useful for determining
 CC the phenotype of a cell by detecting the differential expression of the
 CC sequence relative to a normal cell. The probes derived from the sequence
 CC can also be used to determine the phenotype of cells in a sample. Probes
 CC and antibodies which hybridise to the nucleotide sequence can also be
 CC used to determine the phenotype of a cell. The primers are useful for
 CC detecting a mutation in a test nucleotide sequence and also for detecting
 CC cancer, preferably colon cancer. Antibodies against the protein encoded
 CC by the nucleotide sequence can also be used in a method to detect colon
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
 CC colon cancer at an early stage.
 XX
 SO Sequence 593 BP; 173 A; 131 C; 137 G; 142 T; 10 other;
 Query Match 3.88; Score 123.2; DB 21; Length 593;
 Best Local Similarity 75.68; Pred. No. 4.6e-24;
 Matches 437; Conservative 0; Mismatches 87; Indels 54; Gaps 21;
 QY 2097 actgagacatgcttaggggtgttcttaattcttccctgtgtggaacaagacc 2156
 DB 1 actgagacaaagtta-gggctgttcttcttaattcttccctgtgtggaacaagacc 58
 QY 2157 ccatctcaaatctagaggaagcctcccgacatgcttgcctccctggcgcaaacat 2216
 DB 59 cattctc-aatctagaggaag-ctcccgacatgcttgcctccctggcgcaaacat 111
 QY 2217 gctctctgagttaagtgaacttaactccctggagacacatcgaactgaactgtgag 2276
 DB 112 tgcctctgagttaag-tgacctaatccctggagacata--cgatcaactgtgag 167
 QY 2277 gtccgagggagatgagaaaggaataccacatcttccataggtatcaagaactcactc 2336
 DB 168 gtccgagggg--atgagaaaggatgacccacacatcttcaaggtcacaaagctcacttc 224
 QY 2337 cgtgacagtcagatagaggaacacctgctctatccctccaaaggagagatctggcc 2396
 DB 225 --tgacaagtcagataggg--acactgctctatccctccaaatgga-gagatctggca 279

QY 2397 aaacccccccttlttgaanaaccagggcccccagagcttggcaactagctcaaccag 2456
 DB 280 accctt-----gaacagccaagctt-gaacctagctc-accacag 321
 QY 2457 aagactggaagagacatatctttagcttcttcggggggtgtgctgggaatcag 2516
 DB 322 aagactggaag-gagacataac-ctcagcttcttcagagggctgccc-tgggaatcag 378
 QY 2517 gaactgttgaatcgaatagaaagccttgaactataataatgcatatgggtttc 2576
 DB 379 gaac-tttttgagtcatatagaaagccttgaactataataatgcatatgggtttc 432
 QY 2577 aatctacagtttttgaacatgcttagagagcagaacggggccagagatgaataacatga 2636
 DB 433 actctacagtttttgaac-atgcttagagagcagaagggc-----aaaataaaacatga 486
 QY 2637 cctgttagaaggaaggaagcagaaggaactgggtgg 2674
 DB 487 cctgttagaaggaaggaaggaactgggtgg 524
 RESULT 15
 AAA16148
 ID AAA16148 standard; DNA; 620 BP.
 XX
 AC AAA16148;
 XX
 DT 14-JUN-2000 (first entry)
 XX
 DE Human colon cancer differentially expressed nucleotide sequence #153.
 XX
 KW Colon cancer: detect; differential expression; human; treatment;
 KW detect mutation; non-invasive diagnostic method; ds.
 XX
 OS Homo sapiens.
 XX
 PN WC200012702-A2.
 PD 09-MAR-2000.
 XX
 PF 30-AUG-1999; 99WO-US19424.
 XX
 PR 31-AUG-1998; 98US-0098639.
 PR 27-JAN-1999; 99US-0117393.
 XX
 PA (FARB) BAYER CORP.
 XX
 PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
 PI Catino TJ, Dwiwedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
 PI Schlegel R;
 XX
 DR WPI; 2000-256641/22.
 XX
 PT Novel nucleic acids and proteins for identifying therapeutic agents
 PT useful for treating and diagnosing cancer, especially colon cancer
 XX
 PS Claim 16; Page 185; 345pp; English.
 XX
 CC This sequence represents a human nucleotide sequence which is
 CC differentially expressed in colon cancer cells compared to the expression
 CC levels in normal cells. The nucleotide sequence can be used as a source
 CC of primers and probes. The nucleotide sequence is useful for determining
 CC the phenotype of a cell by detecting the differential expression of the
 CC sequence relative to a normal cell. The probes derived from the sequence
 CC can also be used to determine the phenotype of cells in a sample. Probes
 CC and antibodies which hybridise to the nucleotide sequence can also be
 CC used to determine the phenotype of a cell. The primers are useful for
 CC detecting a mutation in a test nucleotide sequence and also for detecting
 CC cancer, preferably colon cancer. Antibodies against the protein encoded
 CC by the nucleotide sequence can also be used in a method to detect colon
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
 CC colon cancer at an early stage.
 XX

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 13:24:16 ; Search time 8456.85 seconds
(without alignments)
6254.097 Million cell updates/sec

Title: US-09-867-034-4

Perfect score: 3206
Sequence: 1 ttccgctcgcagtgtaaac.....ttgtatgatcgaataaaaaa 3206

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_ov: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vi: *
30: em_htgo_hum: *
31: em_htgo_inv: *
32: em_htgo_rod: *
33: em_htg_hum: *
34: em_htg_inv: *
35: em_htg_rod: *
36: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1466	45.7	2887	9 AF286113	AF286113 Homo sapi
2	1458.8	45.5	2864	9 AB035807	AB035807 Homo sapi
3	1030	32.1	2265	6 AX193124	AX193124 Sequence
4	1030	32.1	2265	6 AK000070	AK000070 Homo sapi
5	1009	31.5	2263	6 AX011684	AX011684 Sequence
6	462.4	14.4	186893	2 AC026342	AC026342 Homo sapi
7	365	11.4	536	6 AX079657	AX079657 Sequence
8	200.2	6.2	417	6 AX192560	AX192560 Sequence
9	187.6	5.9	401	6 AX192474	AX192474 Sequence
10	182	5.7	118407	2 AC026629	AC026629 Homo sapi
11	160.2	5.0	2636	10 RN089744	RN089744
12	149.6	4.7	1326	9 HSM801289	HSM801289
13	149.6	4.7	5182	9 AB023198	AB023198 Homo sapi
14	149.6	4.7	232816	2 AC012362	AC012362 Homo sapi
15	144	4.5	2968	10 MUSANT10A	G21920 human STS W
16	136.8	4.3	294	11 G21920	AC026342 Homo sapi
17	102.6	3.2	186893	2 AC026342	AL592077 Dario rer
18	61.4	1.9	99996	2 AL592077	M16230 Strongyloce
19	60.6	1.9	14091	3 SUSMP1	AJ271723 Fugu rubr
20	60.6	1.9	22398	5 FRU271723	AC020227 Drosophi
21	59.8	1.9	155550	2 AC009749	AC009749 Drosophi
22	59.8	1.9	161278	3 AC009749	AE003608 Drosophi
23	59.8	1.9	241429	3 AE003608	AE003608 Drosophi
24	58	1.8	1470	3 LSI132603	AE003608 Drosophi
25	57.8	1.8	2161	5 XELFTMCLX	L02115 Frog Integu
26	56.2	1.8	79433	2 AL591180	AL591180 Dario rer
27	55.2	1.7	186222	2 AC020858	AC020858 Mus muscu
28	55	1.7	8438	14 AF160185	AF160185 Fowl aden
29	54.6	1.7	379	8 AF286605	AF286605 Pinus tae
30	54.6	1.7	185994	2 AC002042	AC002042 Homo sapi
31	54.6	1.7	327209	3 AE003575	AE003575 Drosophi
32	54.4	1.7	172307	2 AC044842	AC044842 Homo sapi
33	54	1.7	227724	2 AF336381	AF336381 Mus muscu
34	53.8	1.7	64998	2 AC014035	AC014035 Drosophi
35	53.8	1.7	269518	3 AE003751	AE003751 Drosophi
36	53	1.7	5163	6 AR097042	AR097042 Sequence
37	53	1.7	5318	6 AR097041	AR097041 Sequence
38	53	1.7	182978	2 AC021176	AC021176 Homo sapi
39	52.6	1.6	158414	10 AC005960	AC005960 Mus muscu
40	52	1.6	448	11 G45188	G45188 220844.1 ze
41	51.8	1.6	177989	2 AL591430	AL591430 Mus muscu
42	51.8	1.6	200125	2 AL591064	AL591064 Mus muscu
43	51.4	1.6	7218	6 I66494	I66494 Sequence 14
44	51.4	1.6	25666	2 AC020076	AC020076 Drosophi
45	51.4	1.6	323461	3 AE003491	AE003491 Drosophi

ALIGNMENTS

RESULT	1	LOCUS	AF286113	DEFINITION	Homo sapiens transmembrane mucin MUC13 (MUC13) mRNA, complete cds.	VERSION	AF286113.1	GI:14209831	ORGANISM	human.	REFERENCE	1 (bases 1 to 2887)	AUTHORS	Williams,S.J., Wreschner,D.H., Tran,M., Eyre,H.J., Sutherland,G.R. and McGuckin,M.A.	TITLE	MUC13, a Novel Human Cell Surface Mucin Expressed by Epithelial and Hemopoietic Cells	JOURNAL	J. Biol. Chem. 276 (21), 18327-18336 (2001)	PUBMED	11278439	REFERENCE	2 (bases 1 to 2887)	AUTHORS	Williams,S.J., Wreschner,D.H., Tran,M., Eyre,H.J., Sutherland,G.R.
--------	---	-------	----------	------------	--	---------	------------	-------------	----------	--------	-----------	---------------------	---------	--	-------	---	---------	---	--------	----------	-----------	---------------------	---------	--

QY	1789	tattagagatccatcaaatgttggaacccgcgaatggcccccaacatagtataatcat	1848
Db	1561	TATTAGAAAT-CATTAGAAATTTGGAACCCCGCATGGCCCCCAACCAATGTACAGCTATT	1619
QY	1849	tattctagcagttctttagaacaagacatgcabtggaagttagacacaagctaaagactctgg	1908
Db	1620	TTTT-----AGAGTGTTTAAGAAAGACTGATGGAGAAAGTATGAGACCA--GTAAGATCTGG	1671
QY	1909	ccctcggagattctctcttccatcatagacatactgccagtcctcatctgcaatgagcaag	1968
Db	1672	CCCTCCGGGGTTTTTTCTTCATCTGCATGACATCGCCAGCCCTCTGAAATGGAATTTGCAATG	1731
QY	1969	btgtagcaatgtcttgcacaagaaatcccaagctcaactgtgtaaatagaatctatgaat	2028
Db	1732	TTTTGCA-----CGAATTCAGCTCTCACTTGGCT-AAATPAGAAATCTAAGACAT	1767
QY	2029	taaatgttagctcgatgctatagctgcgtgtgcacaagaagtggtgttcttccatcaagt	2088
Db	1777	TA--AAATGTATGTATGATGCTTATAGCCCTTGTGAGAA---GGTGGTTTTTCTTAATCAGT	1831
QY	2089	aacaagactatgagacaatgtcttaagtggtgtgttcttaatcttctccgttagagca	2148
Db	1832	-ACAAATAGTATGAGACAAATGTTA--GGTGTGTTTTCTTAATCTTTT--CCGTGTAGGGCA	1888
QY	2149	acaaagaccacatttccaaatcatagagaagaagctccccaagatctgttctgcctgagc	2208
Db	1889	ACAAAGAACCAATTTCC--AACTGTAGAGAAAG-CTCCCCAGCAATTCCTGTCTCTGGGC--	1943
QY	2209	caaaccaatgcttctttagttaaagttagaacttaatcccttggagaagacataccgataa	2268
Db	1944	--AAACTTTCCTCTTGAGTTAAG-TCACCTAATTTCCCTGGAGACATTA--CGATATA	1997
QY	2269	ctgtgtagtgcgaaggggagatgagaaaggatataccacaatcttcatatgagttcaaac	2328
Db	1998	CTGTGAGGTCGCGAGGGG--ATGAGAAAGGATPACCCACACCTTTCAAGGGTCAACAACCT	2055
QY	2329	tacaactcgtgacaagtcagaatagaagggaacacatgctctatccctccaatgtagaaga	2388
Db	2056	CACCTCT--TGACAACTCATGAATTAAGG--ACACTGCTCTTATCCCTCCCAATGGA-GAGA	2110
QY	2389	tctggagcaaaccccccttlttgsaaacagcccccaagagcttgcgaactgagctc	2448
Db	2110	TTCTGGCAA-----CCTTTGAACAGCCCGAGAGCTT-GCAACTATAGCTCT	2152
QY	2449	aaccacaagagacttggaaaagagagacatacttcaagcttlttcaaggaagcgctgcctgg	2508
Db	2153	-ACCAAGAGAGACTGGAAA-GAACAATATC-TCACAGCTTTTTCAGAGAGCGCTGCC-TGG	2208
QY	2509	gaatccagaagaagttttagtgcataatagaagagcctggaactataataatgcatcat	2568
Db	2209	GAAATCCAGGAAC-TTTTTGATGCTAATTTAGAAAGGCTGAGCTAAAATGTCCAC---TA	2263
QY	2569	ggggtttatactacagttttagaactcttagaagcctggaagcagggcacagagtagaa	2628
Db	2264	TGGGTCACCTCTACAGTTTTTTAA-ATGTAGAGAGCAAGAGGGGACAGAGG--TAA	2319
QY	2629	aaaatataccctgttagaaggaagaagagcaaaagaaacgtggctggggaagatcaataga	2688
Db	2320	AAACATGACCTGGTGAAGAGAAAGAGGCAAGGAAGAACTGGGTGGGAGATCAATTAGA	2739
QY	2689	gaagaagacacctggatccaactgttctcctaagtctccctctccatgacgaagaag	2748
Db	2380	GAGAGGACACCTGGATCCACCTTC-TTCCTTAGTGCCCTCTCCAT-CAGCAAAAGAG	2437
QY	2749	caacttcctatgcatagccctccccaagaagctgcttggggaagaagtttaaaaaaataaa	2808
Db	2438	CACCTTCCTTAA-TCATATCCCTCCCGAAGACTGGCTGGGGAAGGTTT-AAAAACAAAAA	2495
QY	2809	tccagagataaagaccttgaagtgcaattttagaaattgagaaacacttgcctggcaa	2868
Db	2446	TCCGAGAGAT-AAAGGCTTATAGTACAG---TTTGAATTTGGAAACAACAACGT--CTGGCA	2549
QY	2869	agggtgcgaagacggaagcttgttgcacgaagatcccaagcgttcaagcctcggagtgtaa	2928

[illegible]

2232 ttgaacttaactccctctggaagacacatccgcatcaactgtggaaggtccgagggagatga 2291
1962 -TGACCTTAATTCCTCGGAGACATA---CGCATCACTGTGGAGGTCCGAGGGG--ATG 2015
2292 gaaaggataccacacatcttcataagggtacaaagctacactcgtgcaagtcagaa 2351
2016 AGAAGGGATACCCACACACTTTCAAGGGGTCAACAGCTCACTCTC---TGACAAAGTCAGAA 2072
2352 taagggaacactgcttctatccctcccaatggagatgttcggccaacccccctttt 2411
2073 TAGGG--ACACTGCTTCTATCCCTCCAAATGGA--GAGATTCTGGCA----- 2115
2412 tgaaaaccagcccccagagcttgcaacctagcctcaaccacaaagaaagctggaagag 2471
2116 --CCTTTAACAAGCCCAAGCTT--GCACCTAGCCTC--ACCCAGAAACATCTGGAAA--GAG 2170
2472 acatatcttcaagcttcttcagagagcgctgcttgggaatccaggaacgttttgatgc 2531
2171 ACATATC-TCTACACTTTTTCAGAGGCGGTGCC--TGGGAATCCAGGAAC--TTTTGATGC 2227
2532 taattagaagccctgagctatataatgtcatcatatggttttaactatcagttttg 2591
2228 TAATTAGAAGCCTGGAGCTATAAATATGCCAC---TATGGGTGCACTCTACAGTTTGG 2283
2592 aacatgctagaagcaagaacgagcgagcaagagtaaaaaacatgacctgctagaagag 2651
2284 AA-ATGCTAGAGAGGAGAGAGGCGCAGAGAG---TAAAAACATGACCTGTGTAAGAGAG 2339
2652 agagggcaaaagaaactggtggtggaagatcaattagaagagagcaactggtgacact 2711
2340 AGAGGCAAAAGGAAACTGGGTGGGAGATCAATTAGAGAGAGGACCTGGGATCCACT 2399
2712 tctgttcccttagtcccccctccatcgcaagagagcaacttcttaagttaagctccc 2771
2400 TC-TTCCCTTAGTCCCTCCCTCCAT--CAGCAAAAGAGCACTTCTTA--TCAATGCCCTCC 2456
2772 cgaagactgctggaagaggtttaaaaaaacaataaccagagataagagccttagag 2831
2457 CGAAGACTGCGTGGGAGAGGTTT-AAAAACAAAATCCAGAGAT--AAGAGCCTTAGG- 2513
2832 tcaagtttgaataatggaagacaaactgtcttgcaaaaggttgccaagagcgagactgt 2891
2514 --TCAGTTTGAATTTGGAGACAAACT--GTCTGGCAAAAGGCTGC---GAGAGGGAGCTTG 2566
2892 tgcctgaagatgccagccgcttcaagcctcgaggtgtgaagttcttgaaggtgtgccaatg 2951
2567 TGCCTCAGAGATC-CAGCCGCTCAGCCTCGGGGTGT-AGGTTTCTGAGGTGTGCCATTTGG 2624
2952 gctcagccttctctgtgacccgaggtcagctgtgagccacaaacacacacacacac 3011
2625 GCTCAGGCTTCTCTGTGGA--CAGAGGCTCAGCTGTGGCACAACACACACACACAC 2683
3012 acacaaaccacacacaaatggtgggcaaacacacatccacgtaaccaagctttaacacaa 3071
2684 ACACAAACACACACAAATGGGGCAACACATCCACTAACAGC-----TTTTACAAA 2737
3072 tgtattagtgcccttcttcttaataagccctgtcctcttaaaagttaattattg 3131
2738 TGTATTATGTCTCTTTTATTCTAAT--GCTTGTCTCTTAATAAGTATTATTATTG 2796
3132 ttaattatattgtctctgactgttaattgtaatgtaatgcaataaagtgcttct 3191
2797 TTATTATTATTGTCTCTGAGCTTAATTGGAATGTAATGCAATAAAGTGCTTTGTT 2856
3192 agatgg 3197
2857 AGATGG 2862

RESULT 3
AX193124 2265 bp DNA PAT 15-AUG-2001
LOCUS AX193124
DEFINITION Sequence 691 from Patent WO0149716.

ACCESSION AX193124
VERSION AX193124.1 GI:15211077
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2265)
AUTHORS Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,
Stolk,J.A., King,G.E., Wang,Y. and Jiang,Y.
Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use
Patent: WO 0149716-A 691 12-JUL-2001;
JOURNAL CORIAX CORPORATION (US)
FEATURES
source Location/Qualifiers
1..2265
/organism="Homo sapiens"
BASE COUNT 679 a 484 c 522 g 580 t
ORIGIN
Query Match 32.1%; Score 1030; DB 6; Length 2265;
Best Local Similarity 83.2%; Pred. No. 5.3e-252;
Matches 2001; Conservative 0; Mismatches 260; Indels 145; Gaps 67;
OY 801 ttgctctgttgaagaggtatctactcaactcttctcaatgtgaagaagaagrat 860
DB 2 TTGGCTGTGTTTGAAGAGGTAATTAATCAACTCTTCTCAATGTAAAGAGAAAGGTAT 61
OY 861 tcccggagaagatttcagtgacagttatcagaacatttgaccacagagaagaacatcc 920
DB 62 TCCCGG--GAAGATTTCAGTGTACACTATCAGAAACATTGGACCCGAGAGAAACATTCC 120
OY 921 atggcctatcaagactgcatagtgaaatctacagctgtttaaagatglatctggcaca 980
DB 121 ATGGCTATCAAGACTTGCATAGTGAATTAATTAATGAGTGTAAAGATGTAATTTGGCACA 180
OY 981 tctgttcttgacagacgtatcttctactgtgaagacatctcgttaccacaaatcttg 1040
DB 181 TCTGTTATGGAACAACATGTAATCTTACTGTAA--GCACATCTGTGCCAAGA--TCTG 238
OY 1041 aaatcgctgtcttgatgaacatttcttaactgaacaaatgaacaaatcttgacgaag 1100
DB 239 AAATCGGTGC--TGATGACAAG--TTTGTTAATGTAAATTAATTAATTTGGCAGAAAC 296
OY 1101 cacaagtacaatgagaagactgtgactgaggaataatgaagaacattataagtagc 1160
DB 297 CACAAGTGACAATGAGAAGACTGTGACT--GAGAAATTAATTAAGCAA--TTAGAAGTAGC 354
OY 1161 tcaagcaacttcttaactatgatggaacctgtcggtggtggtatgattgaaggctgg 1220
DB 355 TCAAGCAAC--TTTCTAAACTATGATTTGACC-----TTGGGTGATTTATGAGCTG 407
OY 1221 aacaaagactgctggaagactcctcaatggtgttaagcatgcaatgcaatgctac 1280
DB 408 TAACCAAGCTG--CGATGACTGCTCTCAAT--GCTTACATATGCCA--TTGCAAAATCTAC 462
OY 1291 ctgcaagagcctaaccacagaagcccttctgctgtgcttcacagctcgaagtgctcga 1340
DB 463 CTGCAAAAGCCCTAACCCACAGAGCCCTTCTGCGTTGCTTCCAGCTCA--AGTGCTCGA 521
OY 1341 tgcctgcaagcgaagcgaagcgaatgtcttaataaagaagtggttggtccctctga 1400
DB 522 TGCCTGCAACGACACACAAAGC--AATGCTTAATAAAGAGTGTGGGCCCTCGTAGT 580
OY 1401 gttgtcgcttgctgctgctgctacccaagaaatgctaatggaactccaagaagtgtgc 1460
DB 581 GTGGCTGGCTG-----CCGGCTACACAGAGATGCTAATGGAATCTCCAAAAGTGTGC 635
OY 1461 attggactacaagtgactgactgtlaagacaattcaagctgactcactattgt 1520
DB 636 ATT--GGCTACAGTGAGCTGACTGTAAAGACAAATTTACACTGATCTCACT--ATTGT 692

QY	1521	gggcacatgcctgycgatitctcaatcttcagcatgataatgtgatgtgtctactcagaa	1580
Db	693	gggcacacatccctggcatgttcattctcacacatgaataattgcatgtattgtcc-AGCA	751
QY	1581	gatacaatacaaaaagaaagacatattgaaagacgaacactgattgacgaacattcca	1640
Db	752	GATCAATAATACAAA--GGAAGCATATTGAAGA--GAGAACTTGATTGACGAAGACTTTCA	809
QY	1641	aaatctaaactgcgttcgcacagctctcaactcaactctatgtagacatacgaagctctc	1700
Db	810	AAATCTAAATCTGGGGTCG--ACAAGCTTCCACCATCT--TGGAGCGAAGAGGGCTCTTT	867
QY	1701	cctaaggtcagatataagcctcccaagagacgcgcctagattgaaaaatccgtagttca	1760
Db	868	CCTAAGAGTCAGAGATACGGCTCCAGAGA--CAGCCAGATGCANAATCCC-----TATTCA	921
QY	1761	agacacagcagcatgcccgcgcgtgactatagaatcatalcagaatgtgaaacccgca	1820
Db	922	AGACACACACACATG--CCCCGCCCTGACTATTAGAAAT--CATTAAGATTGGAAACCCGCCA	979
QY	1821	tggccccaaccatagtacatatactatactctagcagtggttagaagaactcagtg	1880
Db	980	TT	1033
QY	1881	agaagtgaagccacgtgaagaactctgcccgcggagttctcttcacatagaacata	1940
Db	1034	AGAAGTGAAGACCA--GTAAAGATCTGGCTCGGGGTTTTCTTCCATCTGACATCTGC	1091
QY	1941	ctgcagctcccatctgcgaatgacgaactgtgcgaatgtctctgcaaaagacatccagct	2000
Db	1092	CAGCTCTCTTAATGAAGATTGTGAATGTTTGCA-----CGAATCCACT	1137
QY	2001	cactgtcctaataagatcatctgacataactgtacgtcatctatagcgtctgc	2060
Db	1138	CACTTGCT--AAATAAGATCTATGACATT--AATGTAGATAGTGTATTACCGCTTGTC	1194
QY	2061	tcagagagtggttctctccaatcaglaacaaagtaactgagacaatgtcttaggggttg	2120
Db	1195	AGAGA---GGTGGTTTTCTTCATCACT--ACAAGATACAGACACATGGTA--GGGTGT	1249
QY	2121	ttctctaatctctccctctgtaggaagaagaaccccatctccaatctagagaagac	2180
Db	1250	TTTTCTTAATTTT--CTGTGATGGCCACACAAAC--ATTTCATCTAGAGAAAG-	1305
QY	2181	ctccccaagatctgtctgtccctcctgagccaaacatgctctctgtaagtgcacctaa	2240
Db	1306	CTCCCCACACATTGCT---TGCTCTTGGGCAACATTGCTTGAGTTAG--TGACCTAA	1360
QY	2241	cttcccccggagacacataccgataacactgtgaaagtctccgggggagatgaaagagat	2300
Db	1361	TTCCCCCTGGGACATA---CGCATCAACGTGTGAAGTCCGAGGG--ATGGAAGGAGAT	1415
QY	2301	accacaacatcttcataaggttcaacaagctacacactcgtgaaagatcagaataagagca	2360
Db	1416	ACCACACATCTTTCAAGGGTCAACAGCTCACCTC---TGACAAGTCAGAATATGG--AC	1470
QY	2361	ccgtgtctatccctccaatgtagagagatctgyccaaaaccccccttlttugaacca	2420
Db	1471	ACTGCTTTCTATCCCTCCAAATGA--GAGATCTTGCA-----CCTTTGA	1513
QY	2421	ggccccagagctgtgaaactctgactccaacccaagaagactggaagagacatatct	2480
Db	1514	ACAGCCAGAGCTT--GAACTTAGCTC--ACCCAGAAACTGGANA--GAGACATATC--T	1559
QY	2481	ttcagcttttcagagagcgtctcctggaatccagaaagctlttltgtacaaatagaa	2540
Db	1570	CTCAGCTTTTTCAGAGGCGTGCC--TGGGAATCAGAGAAC--TTTTTGATGCTAATTAGAA	1627
QY	2541	ggcctgagacataataatgtccatctatggggtttaaactacagttlttgaacatgcta	2600
Db	1628	GGCTGTGACTAAATAATCTCAC---TATGGGGTGCACCTCTACAGTTTTTTGAA--ATGCTA	1682

QY	2601	ggaagcagaaaggggccaaagagtagtaaaaaaacatgacccgtgtaaagaagaagagcgaa	2660
Db	1663	GGAGCGAGAAAGGGCGAGAGG---TAAATAACATGACCTGTGTAGAAAGAAAGAGGCAAA	1739
QY	2661	ggaacatgagctggtggagagacaaatltagagagagagcgaccttggatccacacttgcctt	2720
Db	1740	GSAAATCGGGGTGGGAGAGATCATTTAGAGAGGAGGACCTGGGATTCACACTTC---TTCCTT	1798
QY	2721	aggtccctctccatcagcagcaagaagcactctctaaagtaagtcacgtccccgaagactg	2780
Db	1799	AGGTCGCCCTCCTCCAT-CACCAAAAGGAGCACTTCTTTAA-TCATCCCTCCCGAAGACTG	1856
QY	2781	gctgaggaagaggtttaaaaaaacaataaccagagtagaagagccttagtgctcagttgtg	2840
Db	1857	GCTGGGAGAGAGGTTT-AAAAACAAAAAATTCAGAGAGT-AAGAGCCTTAGG---TCAGTTT	1911
QY	2841	aaatltgagagacaaactgtctctgagaaaggtgtccaaagcgcgagactgtgtgtcagga	2900
Db	1912	GAATTTGGAGACAAACT--GTCTGGCAAAAGGGTGC---GAGAGGGAGCGTTGGTGCACGA	1966
QY	2901	gtccagcgcgtccaagcctctggggtgtgaaggtctctgaggtgtgtgcaatggtgggcctcagc	2960
Db	1967	GT-CCAGCGCGTCAGCCTCGGGGGTGT-AGGTTTCTGAGGTGGCCATTTGGGGGCTCACCC	2024
QY	2961	tctctgtgtgacccgagagcctcagctgtgtggccacaacacacacacacacacacacaa	3020
Db	2025	TTTCTGTGTGACA-GAGGCTCAGCTGTGGCGACACACACACACACACACACACACAA	2083
QY	3021	cacacacaaatggtggggaacccacatccacgtcaacccaagcttlaacacaaatgltatag	3080
Db	2084	CACACACAAATGGGGGCAACACATCCAGTACAGC-----TTTTCACAAAGTTATTAG	2137
QY	3081	gtgccctttatcttcaatagccgcgtcctctaaagtaattttattgtttatata	3140
Db	2138	TGTCCCTTTTATTTCGTAAT-GCCTGTGCTCTTAAAGATTAATTTTATTGTATTATTA	2196
QY	3141	tctgtctgacgttbaatgtgaatgtgaatgtcaataaagtgacctgtgtagatgagaa	3200
Db	2197	TTTGTCTGTGACTTAAATTTGTGAATGTGAATGATCATTAAGTGCCCTTTGTTAAAAAAA	2256
QY	3201	aaaaaa 3206	
Db	2257	AAAAAA 2262	

RESULT	4
AK000070	
LOCUS	AK000070 2265 bp mRNA PRI 22-FEB-2000
DEFINITION	Homo sapiens CDNA FLJ20063 f1s, clone COL01524.
ACCESSION	AK000070
VERSION	AK000070.1 GI:7019918
KEYWORDS	Oligo capping; f1s (full insert sequence).
SOURCE	Homo sapiens colon CDNA to mRNA, clone_1lb:COL clone:COL01524.
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (sites)
AUTHORS	Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitama,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S. NEDO human cDNA sequencing project Unpublished (2000) 2 (bases 1 to 2265) Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
JOURNAL	Direct Submission Submitted (15-FEB-2000) to the DDBJ/EMBL/GenBank databases. Sumitomo Sugano, Institute of Medical Science, University of Tokyo, Dep'tme of Virology, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Jap (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
TITLE	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert
COMMENT	

sequencing: Research Association for Biotechnology; cDNA library construction, 5'-6' 3'-end one pass sequencing; Department of Virology and Human Genetic Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES
source

1. .2265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL01524"
/clone_1db="COL"
/issue_type="colon"
/note="Cloning vector pME18SFL3"
121. .954
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA90925.1"
/db_xref="gi:7019919"
/translation="MAYODLSEITSLFKDVGTSVYGQVILFTVSTLSRSEMRAD
DKFVNTVITLAEFTSDNEKTEIKAIKATRSSSNHNLDLRCDYGCNQTAD
CLNGACDDCKSDLRPNOSPECVASLCKGADCAOAKHOKLKKSGAPACACVPG
OEDANGNOKCAFESGLDCKDFOLIITIVGTAGVILSMITLITATSNKTKH
IEEENLIDEDFQNKLNSTGFTNLGAEESVPEPKVITASRDSOMNPYSRHSMPRP
Y"

CDS

BASE COUNT 679 a 484 c 522 g 580 t
ORIGIN

Query Match 32.1%; Score 1030; DB 9; Length 2265;
Best Local Similarity 83.2%; Pred. No. 5.3e-25;
Matches 2001; Conservative 0; Mismatches 260; Indels 145; Gaps 67;

OY 801 ttgttcgtgttagaagggtatctactacaactctctacatgtaagaagaaggtat 860
DB 2 ttgttcgtgttagaagggtatctactacaactctctacatgtaagaagaaggtat 61
OY 861 tccctggaagaattcagtgacatcacaagaacattgaccagaagaagaacattcc 920
DB 62 tccctggaagaattcagtgacatcacaagaacattgaccagaagaagaacattcc 120
OY 921 atgacctcaagaactgcatagtgaaatctactagctgtgttaagaatgattggcaca 980
DB 121 atgacctcaagaactgcatagtgaaatctactagctgtgttaagaatgattggcaca 180
OY 981 tctgtttatgacaacactgtaattctctgtgaaggcacatctgtaaccaagattcgg 1040
DB 181 tctgtttatgacaacactgtaattctctgtgaaggcacatctgtaaccaagattcgg 238
OY 1041 aaatcggtctgtgacaagaattgtgttaattgtaacaatgtaacaatttggcagaac 1100
DB 239 aaatcggtctgtgacaagaattgtgttaattgtaacaatgtaacaatttggcagaac 296
OY 1101 cacaagtgaacaatgagaagaactgtagtgagaanaaataaataagaacattataagtagc 1160
DB 297 cacaagtgaacaatgagaagaactgtagtgagaanaaataaataagaacattataagtagc 354
OY 1161 tcaagaacaatttcaaaactatgtagcctgcgcgtgtgtgattgtgttggaggtcggg 1220
DB 355 tcaagaacaatttcaaaactatgtagcctgcgcgtgtgtgattgtgttggaggtcggg 407
OY 1221 aaccaagaactggtctgtagtgcctcaattggttttagcatgcatgtgcaatgctgac 1280
DB 408 taaacagactg--cggatgactgctcaat--ggtttagatnccga--ttgcaaaacttgac 462
OY 1281 ctgcaagaagcttaaccacagaagccttctggttctgcttccagctcaagtagtctga 1340
DB 463 ctgcaagaagcttaaccacagaagccttctggttctgcttccagctcaagtagtctga 521
OY 1341 tgcctgcaacgacagcagaacgaatcttaataaagaagaatgtgtgtgtccctctga 1400
DB 522 tgcctgcaacgacagcagaacgaatcttaataaagaagaatgtgtgtgtccctctga 580
OY 1401 gtgttcgtgtgctgtccggtctacagaagaatgtctaattggaactgcacaaagtgtgc 1460

DB 581 gtagctggcgtg-----cccggtacacagaaagatcttatgggaactcccaaaatgtgc 635
OY 1461 atttggtctaacgtagtgcactgacgtgtaagaagaacattcgaactgactcctactattgt 1520
DB 636 atttt--gggtacagtgactgcactgaaggacaatttcagctgactgactcctact--attgt 692
OY 1521 gggacacactgctgctgcatgtgcatctcgaagatgtagtaatttgatgtattgttaacga 1580
DB 693 gggacacactgctgctgcatgtgcatctcgaagatgtagtaatttgatgtattgttaacga 751
OY 1581 gatcaataaacaagaacgaagacatactgaagaacgaagaactgattgacgaagaacttca 1640
DB 752 gatcaataaacaagaacgaagacatactgaagaacgaagaactgattgacgaagaacttca 809
OY 1641 aatctaaactgctgctgacagagcttccacacatctctgagcatalaagcagcgtcttc 1700
DB 810 aatctaaactgctgctgacagagcttccacacatctctgagcatalaagcagcgtcttc 867
OY 1701 cctcaggtcagatgtagtgcctcgaagacacgctcagatgtagtaaaatccgtagttca 1760
DB 868 cctcaggtcagatgtagtgcctcgaagacacgctcagatgtagtaaaatccgtagttca 921
OY 1761 agacaacagacatgccccggtcctgactatgatacacaagaatgtgaacccgca 1820
DB 922 agacaacagacatgccccggtcctgactatgatacacaagaatgtgaacccgca 979
OY 1821 tggcccccaacacatgtagatctatctatctagcaggttttagaagaactgcatg 1880
DB 980 tggcccccaacacatgtagatctatctatctagcaggttttagaagaactgcatg 1033
OY 1881 agaagtggcccccaacacatgtagatctatctatctagcaggttttagaagaactgcatg 1940
DB 1034 agaagtggcccccaacacatgtagatctatctatctagcaggttttagaagaactgcatg 1091
OY 1941 ctgcaagtcctcactgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatg 2000
DB 1092 ctgcaagtcctcactgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatg 1137
OY 2001 caactgtctaaataaagaatctatgacatgacatgacatgacatgacatgacatgacatg 2060
DB 1138 caactgtctaaataaagaatctatgacatgacatgacatgacatgacatgacatgacatg 1194
OY 2061 tcaagaaggtggttcttctcaatgacatgacatgacatgacatgacatgacatgacatg 2120
DB 1195 tcaagaaggtggttcttctcaatgacatgacatgacatgacatgacatgacatgacatg 1249
OY 2121 ttcttcaatcttctcctctgtaggcaacaagaacacatcttcaatcttcaatcttcaatctt 2180
DB 1250 ttcttcaatcttctcctctgtaggcaacaagaacacatcttcaatcttcaatcttcaatctt 1305
OY 2181 ctccccagatgctgtgtcctcctggtgccaacacatgcttcttgaagttaagtagcctaa 2240
DB 1306 ctccccagatgctgtgtcctcctggtgccaacacatgcttcttgaagttaagtagcctaa 1360
OY 2241 ctccccagatgctgtgtcctcctggtgccaacacatgcttcttgaagttaagtagcctaa 2300
DB 1361 ctccccagatgctgtgtcctcctggtgccaacacatgcttcttgaagttaagtagcctaa 1415
OY 2301 accccaactcttcaatgtagcacaagcacaagcacaagcacaagcacaagcacaagcaca 2360
DB 1416 accccaactcttcaatgtagcacaagcacaagcacaagcacaagcacaagcacaagcaca 1470
OY 2361 cctgctctatcctcacaatgtagcacaagcacaagcacaagcacaagcacaagcacaagcaca 2420
DB 1471 cctgctctatcctcacaatgtagcacaagcacaagcacaagcacaagcacaagcacaagcaca 1513
OY 2421 ggcctcagagatgtagcacaagcacaagcacaagcacaagcacaagcacaagcacaagcaca 2480
DB 1514 ggcctcagagatgtagcacaagcacaagcacaagcacaagcacaagcacaagcacaagcaca 1569
OY 2481 tcaagcttcttcaagaagcgtgtccttgggaatcagaagcgttcttgaatgtaataa 2540

Db 1570 CTCAGCTTTTTCAGAGCGCTGCC-TGGGAATCCAGGAC-TTTTTCATGCTAATTAGAA 1627
Qy 2541 ggcctgactataataatgcatatcatggttttaactacagttttgaacatgcta 2600
Db 1628 GGCCTGCGATAAATATGTCAC---TATGGGGTGACACTTACAGTTTTCGA-ATGCTA 1682
Qy 2601 ggaagcagaacggggccagagagtaaaaaaacatgacctgtgttagaaggagaagcaca 2660
Db 1683 GGAGGCGAAGGGCGAGAGAG---TAAAAACATGACTGTGTAAAGGAAAGAGGCGCAA 1739
Qy 2661 ggaacatgggtgggagatcatatagagagagcactgtgataccactgttccct 2720
Db 1740 GGAACATGGGGGGGAGAGATATTAGAGAGGACACTGGATCCACTTC-TTCCTT 1798
Qy 2721 agtccctcctccatcagcagaagagcacttcttaagtcacgtccctccgaagactg 2780
Db 1799 AGTCCCTCCTCCAT-CAGCAAGAGGACACTTCTTA-TCATGCCCTCCCGAAGACTG 1856
Qy 2781 gctggagagaggtttaaacaacaaatccagagagtaagaagccttagagtcagtttg 2840
Db 1857 GCTGGAGAGAGTTT-AAAAACAAATAATCCAGAGT-AAAAGCCTTAGG---TCAGTTT 1911
Qy 2841 aaaaatgagaacactgtctgtgcaaaagggtgcaagagcggagcttgtgtctcaga 2900
Db 1912 GAAATTTGAGACAAACT-GTCTGGCAAAAGGGTGC---GAGAGGAGCTTGTGCTCAGGA 1966
Qy 2901 gtcccaagccgtccagcctcgggtgttaaggtctctgagtggtgcacatggggccctcagcc 2960
Db 1967 GT-CCAGCGCGTACGCCCGGGGTGT-AGTGTCTGAGGTGTGCCATTGGGGCCTCAGCC 2024
Qy 2961 ttctctgtgaccggaggtcagctgtgtgtgcccacaacacacacacacacacacac 3020
Db 2025 TTCTCTGTGTGCA-GAGGCTCAGCTGTGGCCACACACACACACACACACACACAC 2083
Qy 3021 cacacaacatggggagagac 3080
Db 2084 CACACACAAATGGGGGAC 2137
Qy 3081 tgcctcttatttcttaataagccctgtcctcttaaaagtatttatttatttattt 3140
Db 2138 TGCTCTGTGTGCAAT-GCCTGTGCTCTTAAAGTATTTATTTATTTATTTATTTA 2196
Qy 3141 ttgtcttgaactgttaattgtgaatgtgaatgaatgaatgaatgaatgaatgaatgaat 3200
Db 2197 TTTGTCTGTGCTGTATGTGTAATGTATGTATGTATGTATGTATGTATGTATGTAT 2256
Qy 3201 aaaaaa 3206
Db 2257 AAAAAA 2262

RESULT 5
AX011684
LOCUS AX011684 2263 bp DNA PAT 06-SEP-2000
DEFINITION Sequence 80 from Patent WO955858.
ACCESSION AX011684
VERSION AX011684.1 GI:9998208
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Schmidt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences obtained from pancreas tumor tissue
JOURNAL Patent: WO 955858-A 80 04-NOV-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BRAND (DE); ROSENTHAL ANDRE (DE); METZGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
source 1..2263
/organism="Homo sapiens"

BASE COUNT 689 a 483 c 520 g 571 t
ORIGIN
Query Match 31.5%; Score 1009; DB 6; Length 2263;
Best Local Similarity 83.0%; Pred. No. 1,2e-246;
Matches 1980; Conservative 0; Mismatches 260; Indels 145; Gaps 67;
Qy 822 attactacaactctctcatcatgtaagaagaagagatctccctggaagatctcagt 881
Db 1 ATTGACAACTCTTCTACATGTAAGAAGAAAGATTCCTCGG-GAAGATTTCAGTG 59
Qy 882 acagatcagaacaatttgaccagaagaagaacatctcatctgacctcaagacttgcat 941
Db 60 ACAGTATCAGAAACATTGTGACCCAGAGAAACATTCCATGGCCTATCAAGACTTCAT 119
Qy 942 agtgaattactagctgtgtttaaagatgtatttgacacatctgtttatgacagactgta 1001
Db 120 AGTGAATTTACTACTTGTTTAAAGATGTATTTGGCAGTCTGTATGTAGCAGACTGTA 179
Qy 1002 attctactgtaagagacatctctgcacaaagtcttgaatgcgtgtctgtatgacaag 1061
Db 180 ATTCTTACTGTAA-GCAGATCTCTGTACCAAGA-TCTGAATTCGTGC-TGATGACAAG 236
Qy 1062 ttctgttaatgtaacaatgtaacaatttggcagaacaacagtgacaatgagaagac 1121
Db 237 -TTTGTATATGTAACTAATAGTAAATTTGGCAGAAACCAAGTGACATGAGAGAC 295
Qy 1122 tgtgactggagaataatgaataaataatgaatgactgtaagcacttcttaacta 1181
Db 296 TGTGACT-GAGAAATTTAATTAAGCA-TTTGAAGTGTAGCTTAAGCAAC-TTCTTAATCTA 352
Qy 1182 tgaatgacccctgtcgtgtgtgattgattgaaggcctggaacaaagactgctgtagac 1241
Db 353 TGATTTACCCCTTC-----GGTGTATTTATTAAGCTGTAAACCAAGCTGC--GGATGAC 404
Qy 1242 tgcctcaatgggtttgaatgcatgcatgtgcaaatgtgcaactgcaaaaggcctaaccacag 1301
Db 405 TGCCCTCAAT-GGTTTACATGCGATGTC--AAATCTGACCTGCAAAAGCCCAACCCACAG 461
Qy 1302 agcccttctgcgtgtgtctcaagctcaagatgctgcaatgctgcaagaacaaagacaa 1361
Db 462 AGCCCTTCTGCGTGTCTTCAGTCTCA-AGTGTCTGATGCTGCGACGACGACGACAA 520
Qy 1362 gcaaatgcttaataaagaagagtggtgtccctgcaagtggtgtgctgtgctgtgctgt 1421
Db 521 GC-AATGCTTAATTAAGAAGAGTGTGGGCCCTGTAGTGTGCTGTGCTG-----CCCGG 574
Qy 1422 ctaccagaagatgcttaatggaactgcaaaagtgtgcatltgggctacagtggactcg 1481
Db 575 CTACACAGGAAGATGCTAATGGGAACTGCCAAAGTGTGCAATTT-GCTTACAGTGAATCG 633
Qy 1482 acgtgaagagcaaatctcaatgcatctcaactatattgttggaacacatgctgcatgt 1541
Db 634 ACTGTAAAGCAAAATTTGAGTGTACT--ATTGTGGGCAACCTTGGCATTGT 691
Qy 1542 catctcagcatgtaattgcatgtatgtgcaatgcaagaatcaataaacaagaagcgaag 1601
Db 692 CATCTTCAGCATGATTAATTCATGTGATTTGCAC-AGCAAAATCAAAATACAAA-CCAAG 749
Qy 1602 cataatgaagaacgagaactgtatgacgaagacttcaaatctaaactgaactgctgtgca 1661
Db 750 CATATTAAGAA-GAGAACTTGATTGACGAAGACTTTCAAAATCTAATAACTGCGGTG-A 807
Qy 1662 caggtctcccaatctatgagatgaacgaagaggtctccctcaggtcgaagatgaagcc 1721
Db 808 CAGCTTTCACCAACT-TTGAGCAAGAGGAGGCTCTTTCTTAAGGTACAGATTAACGGCC 866
Qy 1722 tccaagagacccgtatgatacaaaaatcccgtagttttcaagaacacagcagatgcccgcg 1781
Db 867 TTCA--GAGACAGCCAGATGCAAAATCCC-----TATTCAAGACACAGCAGCATG-CCCCG 919

QY	1782	gccgactatttagatccatcacaagatgtagaacccgcattgcccacacatagtaca	1841
Db	920	CCCTGACTATTAGAT--CATMAGAATGTGGAAACCCGCCATGGCCCCCAACCATGTACAA	978
QY	1842	tatcatattctcagcaggttttagacaagacgcctctgagaagtgagacacagtaag	1901
Db	979	GCTATTATT-----AGAGTGTTTAGAAAGACTGATGGAGAGTGAACCA--GTAAA	10306
QY	1902	acttcgctccggaagttctctctctccatctagacaatactcgcagctccatctgcaat	1961
Db	1031	GATGTGGCCTCCGGGGTTTTTCTTCATCGATGCACATGTGCACACCTCTGTGATGSAATT	10990
QY	1962	ggcaagctgtgcaatgctctcttccaacgacatccacgcctcactgtgctlaaataagaatc	2021
Db	1091	GTGAAATTTTGGCAACG-----AAATCCAGCTCACTTGGCT--AAATMAGAATCT	11359
QY	2022	atgacataacatgtagctcgaatgcatattagagcctgtgctcagagagtgaggtttctc	2081
Db	1136	ATGACATTA--AATGTAGTAGATGATTAAGCCCTGTCAAGA--GGTGGTTTTCTTC	11900
QY	2082	aattacaacaagtlactgagacaatgcttagaggtttgtttctaatctctccctgg	21418
Db	1191	AATACAT--ACAAAGTACTGAGACAATGGTTA--GGTGTGTTTCTTAATCTTTT--CTGG	12474
QY	2142	tagggacaagaagcccatctccaatctagagaaagccctcccaagatgtctgtc	2201
Db	1248	TAGGGCAACAGAAAC--ATTTCATCTTAAGAGAAAG--CTCCCCAGACTTGTCT---TG	13000
QY	2202	cctgtggccaacatgctctcttaagttaagttagtactactccctcctggagcaatcc	2261
Db	1301	CTCTGGGCAAACTTGCTCTTAAGTTAAG--TGACCTAATTTCCCTGGAGACATA---C	13566
QY	2262	gcaccaactgtgagtgccgagggagatagaagaagatataccaacatctctcaaggt	23212
Db	1357	GCATCACTGTGGAGGTCCGAGGG--ATGAGAAGGATATCCACACACTTTCAAGGTC	14144
QY	2322	cacaagctacactctgtgacaagtcagaatagaggacacctgtctatacctccaatg	2381
Db	1415	ACAAGCTCACTCTC--TGACAGTCAAGATAGAG--ACACTGCTTCTATCCCTCAATG	1469
QY	2382	gagagattcttgccaaccccccttttttgaaaacagcccccagagcttggcaac	24411
Db	1470	GA-GAGATTTGCGAA-----CCTTTGACAGCCCGACAGCTT--GCAACC	15111
QY	2442	tagcctaacccaagaagacttgtaagaagacatacttctcaagcttttcagaagagct	2501
Db	1512	TAGCCTC--ACCCAAGAAGATGGAAA--GAGACATATC--TCTAGCTTTTTCAGAGGGCT	15687
QY	2502	gccttgggaatccaaggaagcttcttgatgataatagaagccttgcataataatgtc	2561
Db	1569	GCC-TGGGAATCCAGGAAAC--TTTTGATGCTAATTAGAGGCGCTGGACTAAAATGTCCA	16262
QY	2562	cactatagggttttaatctcaacglttttgaacatgctatgagggcagaacgggccagag	2621
Db	1627	CTAT---GGGGTGCACTCTACAGTTTTTGA--ATCCTTGAGGCGAGAAAGGGCGAGAG	16811
QY	2622	agtaaaaaacaatgacctgttgaagaagaagaaggaagaagaacatcgtgttgaggagctc	2681
Db	1682	---TAAAAAACATGACTCGTTAAGAGAGAGAGGCAAAAGAAACTGGTGGGAGGATC	17381
QY	2682	aattagagaggagcaccttggatccaactctgtctcttaagtcctccctccatctcagcag	27411
Db	1739	AATTAGAGAGAGGACACTGGATTCACACCTTC--TTCTTAGGCCCTCTCTCAT--CAGC	1796
QY	2742	aaaggagacttctcttaagtcacatgcccctccgaagcctgtcgtggagagaagtttaaaaa	2801
Db	1797	AAAGAGACACTTCTCTTA--TCAAGCCCTCCCGAAGACTGGCTGGGAGAGGTTT--AAAA	1854
QY	2802	caaaaaatccagagatgaagagccttagagtgtaagtttaaaattggaacaaactgtc	2861
Db	1855	CAAAAAATCCAGAGAT--AAGAGCCTTAGG---TCACTTTGAATTGTGAGACAAACT--GT	1908
QY	2862	ttggcaaaagtgccaagagcggagctgtgttgctcagaagtcaccagcgtccagcctcg	2921

	RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	6	AC026342/C	Homo sapiens chromosome 3 clone RP11-559L12, WORKING DRAFT SEQUENCE, 24 unordered pieces.	AC026342	AC026342.18	GI:11995496	HTGS_PHASE1; HTGS_DRAFT.	human.
								Homo sapiens
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS			Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
			1 (bases 1 to 186893)					
			Muny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,					
			Alsbrooks, S.L., Amaralunge, H.C., Are, J.B., Banks, T., Barbara, J.,					
			Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J.,					
			Boile, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Butay, C.,					
			Burke, P., Burkelt, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,					
			Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,					
			Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,					
			Coyler, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,					
			Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Deigard, O.,					
			Demm, A.L., Ding, Y., Dinh, H.H., Douthevalte, K.J., Draper, H.,					
			Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C.,					
			Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,					
			Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,					
			Gatza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,					
			Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A.,					
			Hernandez, J., Hernandez, O., Hodgson, A., Hughes, M., Holloway, C.,					
			Hollins, B., Homs, F., Howard, S., Huber, J., Hulys, S., Humé, J.,					
			Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jojivee, S.,					
			Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvath, J.,					
			Kovar, C., Kratochovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,					
			Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,					
			Loisged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,					
			Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,					
			Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,					
			Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,					
			Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,					
			Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S.,					
			Oguchi, M., Okwonou, G., Ofagunye, N., Oyedero, R., Pace, A., Payton, B.,					
			Pearcy, J., Perez, L., Peters, L., Pickens, R., Plinius, E., Pu, L.L.,					
			Ruiz, M., Ren, Y., Rivers, M., Rojas, A., Rojiboken, I., Rolfe, M.,					
			Savery, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N.,					
			Sisson, I., Sodegren, E., Sonalke, T., Sparks, A., Stanley, H.,					
			Stone, H., Sutton, A., Svatek, A., Tabot, P., Tamerisa, A., Tamerisa, K.,					
			Tang, H., Tansey, J., Taylor, C., Taylor, T., Teifrod, B., Thomas, N.,					
			Thomas, S., Usmani, K., Vasquez, L., Warren, V., Villalon, D., Vinson, R.,					
			Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,					

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8653
Center Clone name: 816_1_3

* NOTE: This record contains 141 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
758 857: gap of 100 bp
858 1584: contig of 727 bp in length
1585 1684: gap of 100 bp
1685 2449: contig of 765 bp in length
2450 2549: gap of 100 bp
2550 3306: contig of 757 bp in length
3307 3406: gap of 100 bp
3407 4163: contig of 757 bp in length
4164 4263: gap of 100 bp
4264 5024: contig of 761 bp in length
5025 5124: gap of 100 bp
5125 5878: contig of 754 bp in length
5879 5978: gap of 100 bp
5979 6734: contig of 756 bp in length
6735 6834: gap of 100 bp
6835 7583: contig of 749 bp in length
7584 7683: gap of 100 bp
7684 8446: contig of 763 bp in length
8447 8546: gap of 100 bp
8547 9304: contig of 758 bp in length
9305 9404: gap of 100 bp
9405 10161: contig of 757 bp in length
10162 10261: gap of 100 bp
10262 11003: contig of 742 bp in length
11004 11103: gap of 100 bp
11104 11831: contig of 728 bp in length
11832 11931: gap of 100 bp
11932 12694: contig of 763 bp in length
12695 12794: gap of 100 bp
12795 13547: contig of 753 bp in length
13548 13647: gap of 100 bp
13648 14410: contig of 763 bp in length
14411 14510: gap of 100 bp
14511 15268: contig of 758 bp in length
15269 15368: gap of 100 bp
15369 16119: contig of 751 bp in length
16120 16219: gap of 100 bp
16220 16986: contig of 767 bp in length
16987 17086: gap of 100 bp
17087 17816: contig of 730 bp in length
17817 17916: gap of 100 bp
17917 18663: contig of 747 bp in length
18664 18763: gap of 100 bp
18764 19505: contig of 742 bp in length
19506 19605: gap of 100 bp
19606 20362: contig of 757 bp in length
20363 20462: gap of 100 bp
20463 21180: contig of 718 bp in length
21181 21280: gap of 100 bp
21281 22041: contig of 761 bp in length
22042 22141: gap of 100 bp
22142 22903: contig of 762 bp in length
22904 23003: gap of 100 bp
23004 23764: contig of 761 bp in length
23765 23864: gap of 100 bp

23865 24598: contig of 734 bp in length
24599 24698: gap of 100 bp
24699 25463: contig of 771 bp in length
25470 25569: gap of 100 bp
25570 26322: contig of 753 bp in length
26323 26422: gap of 100 bp
26423 27173: contig of 751 bp in length
27174 27273: gap of 100 bp
27274 28034: contig of 761 bp in length
28035 28134: gap of 100 bp
28135 28876: contig of 742 bp in length
28877 28976: gap of 100 bp
28977 29730: contig of 754 bp in length
29731 29830: gap of 100 bp
29831 30587: contig of 757 bp in length
30588 30687: gap of 100 bp
30688 31437: contig of 750 bp in length
31438 31537: gap of 100 bp
31538 32297: contig of 760 bp in length
32298 32397: gap of 100 bp
32398 33152: contig of 755 bp in length
33153 33252: gap of 100 bp
33253 34009: contig of 757 bp in length
34010 34109: gap of 100 bp
34110 34856: contig of 747 bp in length
34857 34956: gap of 100 bp
34957 35714: contig of 758 bp in length
35715 35814: gap of 100 bp
35815 36564: contig of 750 bp in length
36565 36664: gap of 100 bp
36665 37401: contig of 737 bp in length
37402 37501: gap of 100 bp
37502 38273: contig of 772 bp in length
38274 38373: gap of 100 bp
38374 39132: contig of 759 bp in length
39133 39232: gap of 100 bp
39233 39996: contig of 764 bp in length
39997 40096: gap of 100 bp
40097 40858: contig of 762 bp in length
40859 40958: gap of 100 bp
40959 41726: contig of 768 bp in length
41727 41826: gap of 100 bp
41827 42591: contig of 765 bp in length
42592 42691: gap of 100 bp
42692 43453: contig of 762 bp in length
43454 43553: gap of 100 bp
43554 44307: contig of 754 bp in length
44308 44407: gap of 100 bp
44408 45157: contig of 750 bp in length
45158 45257: gap of 100 bp
45258 45977: contig of 720 bp in length
45978 46077: gap of 100 bp
46078 46841: contig of 764 bp in length
46842 46941: gap of 100 bp
46942 47693: contig of 752 bp in length
47694 47793: gap of 100 bp
47794 48538: contig of 746 bp in length
48539 48639: gap of 100 bp
48640 49387: contig of 748 bp in length
49388 49487: gap of 100 bp
49488 50243: contig of 756 bp in length
50244 50343: gap of 100 bp
50344 51101: contig of 758 bp in length
51102 51201: gap of 100 bp
51202 51960: contig of 759 bp in length
51961 52060: gap of 100 bp
52061 52818: contig of 758 bp in length
52819 52918: gap of 100 bp
52919 53662: contig of 744 bp in length
53663 53762: gap of 100 bp
53763 54516: contig of 754 bp in length
54517 54616: gap of 100 bp
54617 55380: contig of 764 bp in length

```
* 55381 55480: gap of 100 bp
* 55481 56232: contig of 752 bp in length
* 56233 56332: gap of 100 bp
* 56333 57099: contig of 767 bp in length
* 57100 57199: gap of 100 bp
* 57200 57955: contig of 756 bp in length
* 57956 58055: gap of 100 bp
* 58056 58808: contig of 753 bp in length

Query Match
Best Local Similarity 67.2%; Pred. No. 2.2e-35;
Matches 556; Conservative 0; Mismatches 213; Indels 58; Gaps 25;

QY 2010 aaataagaatctatgacatctagctatgctatctatgagcgtgctgctgagagag 2069
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 111958 AATTAAGAACTATGACATTA--AATGTAGTAGGGCTATAGCGCTAGCAGAGA---G 112012

QY 2070 tgggttcttcttaatactgaataaagtactgagacaatgcttaggggttggttctta 2129
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112013 GTGGTTTCTTCAATCAGT-ACAAGTACTGAGACAAATGGTTA-GGGTTGTTTCTTAAT 112070

QY 2130 tcttccctgtaggagcaaacaccatcttccaaatctagaaggaagccctccagc 2189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112071 TCTTTT-CCGCTAGGGGCAACAAGAAC--ATTTCATCTAGAGGAAAG-CTCCCGACG 112126

QY 2190 attgcttgcctccctggycccaaccatgcttcttgaagttgaagttgaacttccctg 2249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112127 ATTGCT---TGCTCTGGGCAACAACTGCTCTGAGTTAAG-TGACCTAATTCCTCG 112181

QY 2250 ggaacgacataccgacatcaactgtggaagtcgaggggagatgagaagaagataccacat 2309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112182 GAGACATTA---CCGATCAACTGTGGAGGCTCGAGGG--ATGGAAGGATACCCACCAT 112236

QY 2310 ctctcaataggctacaagcactcctgctgacaagctacaatagagagacacctgctct 2369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112237 CTTTCAAGGCTCAACAGCTCACCTC---TGACAAGTCACAATAGG--ACACTGCTTCT 112291

QY 2370 atccctcaatgagagagatcttgcgcaaaccccccttcttgaacaacagagcccccag 2429
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112292 ATCCCTCAATGGA-GAGATTCTGCGAA-----CCTTGACACAGCCGAG 112334

QY 2430 agcttggcaacctagctcaaccagaagactggaagagagacatcttctcagctt 2489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112335 AGCTT-GCAACCTAGCCTC-ACCCAGAAGACATGGAAA-GAGACATATTC-TCTAGCTTT 112390

QY 2490 ttcagaagcgctgctcctggaatccaggaagcttcttgaatcgaatagaagcgctgac 2549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112391 TTCAGGAGGCGTGCC-TGGGAATCCAGGAAC-TTTTGTAGCTAATTAGAAGGCTTGAC 112448

QY 2550 tataatagtlcatctatggtltaatactacagtlttgaacatgactagaagcgaga 2609
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112449 TAAATAATGTCAC---TATGGGGTGACACTCTACAGTTTGTGAA-ATGCTANAGAGCAGA 112503

QY 2610 acggggccgagagagtaaaaaatacagcgtgtagaagagagagagagcaagaaaactg 2669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112504 AGGGGCAAGAG---TAAATAAATGACCTGCTGTAAGAGAGAGAGCAAGGAAACTGG 112560

QY 2670 gt-ggggaagatcaatctagaagagagagcactgcatccactcgttctccttggctccc 2728
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112561 GTGGGGAGAGATTAATTAAGAGAGAGCAACNTGGGATCCACCTTCTCTTATGNNNNNN 112620

QY 2729 tctcctacgagcaagaagagcacttcttaagtcatgcccctccgaagactgagcgag 2788
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112621 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 112680

QY 2789 aaggtttaaataaaaaataccagagatgaagaagccttgggtcag 2835
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112681 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 112727

RESULT 11
RN089744 RN089744 2636 bp mRNA ROD 19-MAR-1997
```

```
DEFINITION Rattus norvegicus putative cell surface antigen mRNA, complete cds.
ACCESSION U89744
VERSION U89744.1 GI:1890274
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2636)
AUTHORS Hajdu, A. and Planagan, P. R.
TITLES Direct Submission
JOURNAL Submitted (18-FEB-1997) Biochemistry, University of Western
Ontario, London, Ontario, Canada
FEATURES
source
1..2636
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/lisue_type="Intestine"
/clone="19-1"
17..1660
/note="similar to murine cell surface antigen 114/A10,
Swiss-Prot Accession Number P19467"
/product="putative cell surface antigen"
/db_xref="GI:1890275"
/translation="MSQSSGGTSPPTTAOPTSTSTONPGTOLLSTSPPTTATQ
PSTSTONPGTTOLOPSTSTONPGTTOLOPSTSTONPGTTOLOPSTSTONPGTTOLOPST
SFQPTRETTLOPSTSTPTTATOPTSTASOTGTTOPPGASSPTTIVYOPGSSQT
PCTOPPGASTPTTIVTOPGSSSOTSGTTPPGGASSTVTSSTSDNDNSNP
CKSPACVRLYDSYFCLLEGYYNNSSCVGTGTPGIGASVETDLESDKSNVY
OTLHSSVVFKEFTKDYGOVILKYKSDLSMSRSYVRAATOTVYVYVNMGEN
TKDEBSVASVKEAVKTDNNYERFOORCYGCKVSGSNCRGLQCTKPGIER
LNDPVEFCVAPKSEPCSAEKKOCLKNGAMEGCKYRKANGKECECEPGYSGM
DKDPOQLITVYGTAGAFILILITVYVMSRSKRYKSGEGEONLIEDPFINLRNP
TGSNAGADTSIFPKVKTGVPSTQSPNPMHMSMPD"
BASE COUNT 734 a 711 c 605 g 579 t 7 others
ORIGIN

Query Match
Best Local Similarity 58.6%; Score 160.2; DB 10; Length 2636;
Matches 554; Conservative 0; Mismatches 343; Indels 48; Gaps 14;

QY 788 tataatacaagtttctgacctgtttagaagggatattactacaac--tctctacatgt 844
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 701 TATGACAGCTACTTCTGCTGTGTTTGAAGAGCTACTACTACAACTTCTTGCTATGT 760

QY 845 aagaagaagaglatccctgagaagatctcagtagacagtalcaagaacatttgacc 904
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 761 GTGAAGAAGAGACGACATCTCCCTGGAG-AGATTGGCATGAGCGTAAAGTAAACAACACTGACTT 819

QY 905 agaagaagaacatctcatagcctcatataagactgcatagagaattactactgtgtttaa 964
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 820 GGAAGATAAATACTCTGTGACTATCAACTTGTGCACAGCAAGTGTTGTTAAGTTTGTGA 879

QY 965 agatgtatctggacatctgtttagacagagactglaactcttactcttgaagacatctc 1024
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 880 AAACACGTTTAAAGAAGACTAGTACAGACTGTGATTTCTTAAAGTAGCAAGCAAGCTC 939

QY 1025 tgtcaccagaatctgaaatgctgtctgagtg--acaagttgtttaaagtaacaagtg 1082
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 940 TTTTAATGTCATCAAGATCTGTGATGCGTCTGCTACGACAGACAGCTATATACAGTGT 999

QY 1083 aacaatttggagcaaaccaagaatgacaatgagaagactgtgactggygaaataata 1142
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1000 AATATATGTTTGGAAACCAAGAGAGATGAAGAGAGTGTAGCTTCTG-TGATCAAGG 1058

QY 1143 aagcaatttaagtagtcataagcaacttcttaactatgatgtgacacctgctgctgtg 1202
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1059 AGGCAGTTAAGACGATACCAATGTGAGACGCTA-----TTTCCACAAGATCG 1107
```

```

QY 1203 gattgatgagggcgtggaaccaagacgtgctgtagtgcctcaatggtttgacatg 1262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1108 GTGTGATTATATATGGCTGTGTAAGAGGTGGACGAATGTTTGGCGAATGGTTTCAATMG 1167
QY 1263 cgaatgtcaaatgctgacgtcgaagaagcctaacccagaagaccccttctgctgtcc 1322
    || || || || || || || || || || || || || || || || || || || || ||
Db 1168 C-ACGTGCAAACTGGGCTGTGAGAGTGAACCCGACAGTTCCTTTTGTGTCTCT--- 1222
QY 1323 agtctcagaagtgctgtcgtcgtcgaagcgaagcagaagc--gagatcctaataaaga 1380
    || || || || || || || || || || || || || || || || || || || || ||
Db 1223 ----CCAAACATGTTCTGAGCCCTGCAGTGCAGAGAAAGACAGCTTATGCTTTAAAGAGA 1278
QY 1381 gagtgytgygggtccctcagtgctgtgctgtgcccgtctacagaagatgctaat 1440
    || || || || || || || || || || || || || || || || || || || || ||
Db 1279 TAATGTAGACGATGAGAGTG-----TGGGTGCATGGAGGCTTACCGGAGGCGAAGC 1327
QY 1441 ggggaactccaaagatgtgcatcttggtgctacaagtgactgactgttaaggacaaat 1500
    || || || || || || || || || || || || || || || || || || || || ||
Db 1328 GGGAAATGTGAAGAGTGT-CCTTTTGGCTACAGCGGATGATGCAAAAGACCAATTTCA 1386
QY 1501 gctatcctcctattatgtgtggaacacatgcctgacatgtcatctcctaagatgtaatt 1560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1387 GCTGATCTCTCAACA--TTGTGGAAACATGTGCTGGAGCTTTCATCTCTGCTAAAT 1444
QY 1561 gcatgtatgtcactagaacatcaataacaagaagcgaagcatatgaaagaacagagac 1620
    || || || || || || || || || || || || || || || || || || || || ||
Db 1445 GTTTTATTCGTCCTCATGAGTGCAGAAAGAAAGAAAGAGAGTGA---GAAGAGCAGAGAT 1501
QY 1621 ttgattgagaagacttcaaatcctaaactgcgtgctgacagaagcttcacaaatctatg 1680
    ||||| ||||| ||||| ||||| || || || || || || || || || || || || ||
Db 1502 CTGATTTGAGATGATCTTCAATACCTGCGAATGAGG-CCAAACGCGCTCTCCAA-CTTCG 1559
QY 1681 gagataacgagagcgtctcctcctcagtcagatgaagctccca 1725
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1560 GAGCAGACACCAACATCTTCCCAAGTCAAAACAGGGGTCCTCGA 1604

```

```

RESULT 12
LOCUS HSM801289 1326 bp mRNA PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; CDNA DKFZP586J1521 (from clone DKFZP586J1521).
ACCESSION AL122124
VERSION AL122124.1 GI:6102978
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1326)
Ostenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-OCT-1999) MIPS, Am Klopfersplitz 18a, D-82152
Martinsried, GERMANY
COMMENT
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Medigenomix (Martinsried/Germany) within the CDNA
sequencing consortium of the German Genome Project. This clone
(DKFZ586J1521) is available at the RZPD in Berlin. Please contact
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.

```

```

FEATURES
Source
1..1326
Location/Qualifiers
1..1326
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP586J1521"
/clone_lib="586 (synonym: hute1). Vector pSport1; host
DH10B, sites NotI, SalI/MuiI"
/dev_stage="adult"
/tissue_type="uterus"
polyA_signal
1299..1304

```

```

polyA_site 1307
BASE COUNT 432 a 190 c 220 g 484 t
ORIGIN
Query Match 4.7%; Score 149.6; DB 9; Length 1326;
Best Local Similarity 97.4%; Pred. No. 3.1e-27;
Matches 152; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 5 gctcagtgtaaaactgcgaagaagtaatactctgtagaggttctgagctgaaga 64
    || || || || || || || || || || || || || || || || || || || || ||
Db 790 GATCCTGTGTAACCTGCGCAAGAAAGTAATTAACCTGTAAGAGTTCGTGACCTTGAAGA 849
QY 65 gtgaactgtgtgtaagagcctgatacaaaaggaagcgaagcattcatattccca 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 850 GTGAATACTGTTGTAATGAGCCTTATCATTAACGAGACCAAGCCATTAATTCCTCA 909
QY 125 agtctaatactgactatgcaatcaacaa 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 910 AGTGTATATATCTGACTTATGCAATTCACAAACA 945

```

```

RESULT 13
LOCUS AB023198 5182 bp mRNA PRI 16-JUN-1999
DEFINITION Homo sapiens mRNA for KIAA0981 protein, partial cds.
ACCESSION AB023198
VERSION AB023198.1 GI:4589605
KEYWORDS
SOURCE Homo sapiens adult male brain cDNA to mRNA, clone_11d:pbuescript11
SK plus clone:hj07094.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
1 (sites)
Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirosewa,M.,
Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
Prediction of the coding sequences of unidentified human genes.
XIII. The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro
DNA Res. 6 (1), 63-70 (1999)
JOURNAL MEDLINE 99246063
REFERENCE
2 (bases 1 to 5182)
Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-1999) to the DDBJ/EMBL/GenBank databases. Osamu
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdna1nfo@kazusa.or.jp; Tel: +81-438-52-3913,
Fax: +81-438-52-3914)

```

```

FEATURES
Source
1..5182
Location/Qualifiers
1..5182
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hj07094"
/clone_lib="h07094"
/dev_stage="adult"
/sex="male"
/tissue_type="brain"
1..1738
/gene="KIAA0981"
<1..1738
/gene="KIAA0981"
/codon_start=2
/product="KIAA0981 protein"
/protein_id="BAA76825.1"
/db_xref="GI:4589606"
/translation="PSVPPSPGRILRQGESEKISAMDASPRNISPGLONGEKEDREFLT
LSOSSSTSTHLQIPPEPVMSBDSVSGRPEDLTASSSEDEVDHILGSDSYKES
TKRALEIPANLGNSTNPLPPFDKHLMEHREVRVFAVEKEPSSITAPALUCKE
RNALEIETSKATQWNSAEGLPTNSTDSRPSSPILPEVSGGTNRRTETEQPK
KASGMLSEFRGTACKSPDISQKRETLTGADSAVAYGVGTGEGTENQGVPEQDEVDG
GDTQKOLINPHEIQFSDANAKFYCRILYAGFPHKKREVILDSSEDFIRLSHSSP

```

MWARGKSGAATFATEDDREIFLKMPRELVOSFLDPAPIHFNITNAVOCKRTALAK
 ILGVNRYIGKNSQNTPEKRLDLVMELEFGKMAOVFLDKSLRRNRKTDGKESC
 DVLADENLKLKRNPNLEKIRSHSKAVLEFTRSHSDSHFLSHLIIIDYSLVGRDPTSN
 ELVVGIIIDYIRFPTMDKLEMYVKSTGILGGQCKMPTVSPELYRRFRFCBAMKYFLM
 VPDHMTGLGINC"

BASE COUNT 1606 a 888 c 1045 g 1643 t
 ORIGIN

Query Match 4.7%; Score 149.6; DB 9; Length 5182;
 Best Local Similarity 97.4%; Pred. No. 3.4e-27;
 Matches 152; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 5 gctcgagtgtaaaactgccaagaagtaattactgttagagttgtcgtgagcttgaga 64
 Db 4653 GATCCTGTGTAAACCTGCCAAGAACTAATTACCTGTAGAGTTGCTGAGCTTGAAGA 4712

Oy 65 gtgaaaaactgttgtaatgagctgatacaaaaagcagccatcatctatctca 124
 Db 4713 GTGAAAACCTGTGTGAATGAGCCGTGATCATPAAACGGACGCAATTCATTCTCTCA 4772

Oy 125 agtctaatactactgtatgcagatccaaca 160
 Db 4773 AGTCTAATATACTGACTTATGCAGATTTCACAACA 4808

RESULT 14
 AC012362 232816 bp DNA HTG 30-JUL-2001
 LOCUS Homo sapiens chromosome 2 clone RP11-404D15, WORKING DRAFT
 DEFINITION
 AC012362 SEQUENCE, 1 unordered pieces.
 AC012362 AC012362.6 GI:15029477
 VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 232816)
 AUTHORS Waterston,R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 232816)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (25-OCT-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 COMMENT On Jul 30, 2001 this sequence version replaced gi:14572170.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: MUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Project Information -----
 Center project name: H_NH0404D15
 ----- Summary Statistics -----
 Sequencing vector: M13; 368
 Sequencing vector: plasmid; 648
 Chemistry: Dye-primer ET; 368 of reads
 Chemistry: Dye-terminator Big Dye; 648 of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 232390 bases at least Q40
 Consensus quality: 232626 bases at least Q30
 Consensus quality: 232815 bases at least Q20
 Insert size: 226000; agarose-fp
 Insert size: 232816; sum-of-contigs
 Quality coverage: 8.23 in Q20 bases; agarose-fp
 Quality coverage: 8.70 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 232816: contig of 232816 bp in length.
 Location/Qualifiers
 1. 232816
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone="RP11-404D15"
 1. 232816
 /note="assembly_name:Contig5
 clone_end:SP6
 vector_side:right"

misc_feature

BASE COUNT 67913 a 42697 c 46345 g 75861 t
 ORIGIN

Query Match 4.7%; Score 149.6; DB 2; Length 232816;
 Best Local Similarity 97.4%; Pred. No. 4.2e-27;
 Matches 152; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 5 gctcgagtgtaaaactgccaagaagtaattactgttagagttgtcgtgagcttgaga 64
 Db 110309 GATCCTGTGTAAACCTGCCAAGAACTAATTACCTGTAGAGTTGCTGAGCTTGAAGA 110368

Oy 65 gtgaaaaactgttgtaatgagctgatacaaaaagcagccatcatctatctca 124
 Db 110369 GTGAAAACCTGTGTGAATGAGCCGTGATCATPAAACGGACGCAATTCATTCTCTCA 110428

Oy 125 agtctaatactactgtatgcagatccaaca 160
 Db 110429 AGTCTAATATACTGACTTATGCAGATTTCACAACA 110464

RESULT 15
 MUSANT10A 2968 bp mRNA ROD 12-JUN-1993
 LOCUS Mouse cell surface antigen 114/A10 mRNA, complete cds.
 DEFINITION
 J04634 J04634.1 GI:191943
 VERSION
 KEYWORDS cell surface antigen.
 SOURCE Mouse (strain C57B16) adult bone marrow hemopoietic cell line
 B6StuA, cDNA to mRNA, clones CDM8.A10.[1,2].
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2968)
 AUTHORS Dougherty,G.J., Kay,R.J. and Humphries,R.K.
 TITLE Molecular cloning of 114/A10, a cell surface antigen containing
 highly conserved repeated elements, which is expressed by murine
 hemopoietic progenitor cells and interleukin-3-dependent cell lines
 J. Biol. Chem. 264, 6509-6514 (1989)

JOURNAL
 MEDLINE
 COMMENT Draft entry and computer-readable sequence for [1] kindly submitted
 by R.K. Humphries, 24-MAR-1989.
 Location/Qualifiers

FEATURES
 source
 1. 2968
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 <1. 2111
 /note="cell surface antigen mRNA (alt.)"
 <1. 2162
 /note="cell surface antigen mRNA (alt.)"
 <1. 2268
 /note="cell surface antigen mRNA (alt.)"
 <1. 245
 /note="cell surface antigen 114/A10 signal peptide"
 195..1916
 /note="cell surface antigen 114/A10 precursor"

```
/codon_start=1
/protein_id="AAA37239.1"
/db_xref="GI:309106"
/translation="MKGFLLISLLILVTVSSSSQASSTSSSGCTSPPTVQSQSP
SSSQASTTSSSGASPPPTVQSQSPSSSQASTTSSSGASPPPTVQSQSPSSSQ
ASTTSSSGASPPPTVQSQSPSSSQASTTSSSGASPPPTVQSQSPSSSQASTT
TSSSGASPPPTVQSQSPSSSQASTTSSSGASPPPTVQSQSPSSSQASTT
GASSIVPSSGSGTSPDLNPNPCIKSKHCLCEGYXNSLSQVKG
TFPEDI SMSVETANLEDENSVGYOELNVTDEFETPNKTDYGTIIVKVSAPSR
SARSAMDARKDVSYSVNTFGADTKETKVSATIEATITSNVADYINICDH
GCVGNDSKCODILIOCTCKPOLDLNPOVPCVAVTCSQPCNAEKEQCLMDGVM
CVCMPGYORANGNRCECPFGYSGMNCKDQPOLILIVGIAIALILILIAIVSA
RSKNKKRDGEORLIEDDFHNLRLRQTFGSMIGADNSIFPKVFGVPSQTPNPYANOR
SMRPDY"
repeat_region
243..883
/note="Ser-Thr tandem repeat region"
mat_peptide
246..1913
/note="cell surface antigen 114/A10"
polya_signal
2091..2097
polya_signal
2951..2956
BASE COUNT      783 a      756 c      707 g      722 t
ORIGIN
```

```
Query Match      4.5%; Score 144; DB 10; Length 2968;
Best Local Similarity 59.2%; Pred. No. 8.8e-26;
Matches 568; Conservative 0; Mismatches 345; Indels 47; Gaps 17;
```

```
QY 789 ataatacaagttttgctcgtgtttagaagggtactactacaactctcta-----cat 842
Db 949 ATAGCAAAACACTCTGCTGTGTGTAGAAAGCTATTACTACAACTGCTCTGTGCTCAT 1008
QY 843 gtaagaagaagaagtttccctcgtgagaagatttcaagtaacagttacgaacatttgac 902
Db 1009 GTGTGAAGAGAACGACATTTCCCTGGGACATTAGCA-TGAGTGTGTCTGAACAGCCAAC 1067
QY 903 ccagaagaagaacattccatgacctcaagaactgtcatagtgaaattactagctgtt 962
Db 1068 TTGGAAGATGAACACTCTGTGGGCTATCAAGAGTGTACAACTGTTACCGATTTTTT 1127
QY 963 aaagatgtatttggcacatcgtttagaacagactgttaattctactgt-----aagc 1017
Db 1128 GAAACACATTTAATAAGACTGACTGACAGACTGTAATTAATAAGTGACAGACGCC 1187
QY 1018 acatctcgtcacaagaattcgaatcgtgctgtgatacagaagtttgttaattgata 1077
Db 1188 CTTTCACGCTCAGCAGAG-TCTGCATATGCTGAT--GCTACAAAGCAGCTCAGTATCC 1244
QY 1078 atagtaacaatttggcagaacaacacaagtgacaatgagaagactgtgactgagaat 1137
Db 1245 GTAGTAATATATTGGAGCAGACACAAAGAACTGAAGAGTATCTTCTGCATTC 1304
QY 1245 gtagtaaatatatttggagcagacacaaagaaactgaaagagtgatcttctgcata 1304
Db 1305 gaa-----ACTGCATTTAAGACTGATGGAAGACTATGTCAGTAAATTTG 1358
QY 1305 taataaagaacattataagtgactcaagaacatttctaactatgattggaacctgtcg 1197
Db 1358 gaa-----ACTGCATTTAAGACTGATGGAAGACTATGTCAGTAAATTTG 1358
QY 1198 gtgtgtgattgttagaggtcgtggaaccaagactgtgtgtagactgctcaatggttta 1257
Db 1359 TGTGATCATTTATGTGTGTGTAGGGATGACAGCATTAATGCGCAGACATTTTACAAATGC 1418
QY 1258 gcatgcatgtgcaaatctgaacctgaacgaagcccaaccagaagcccttctcgttg 1317
Db 1419 ACGTGC-----AAACCTGGGCTGGACAGGCTGAACCCACAGGTCCTTTTGTGTG 1470
QY 1318 ctccagctcagaagtgctcgtgactgtcaacgacacagaacgaagcgaatgcttaataa 1377
Db 1471 CT-----GTAACGTCTCTCAGCCCTGCAATGACAGAGAGAGAG-GAGCAGTCTTAA 1522
QY 1378 gaagaagtgtgtgtccctcgaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1437
Db 1523 GATGATTAACGAGATTAATGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1580
QY 1438 aatgggaactgccaaagtgtgacttgggtacagtgactgcactgtaagacaatt 1497
```

```
Db 1581 AACAGAAAGTGTGAAGAGTGTCTCATTT--GGCTACAGCGGAGATGATTCGAAAGACCAATT 1639
QY 1498 tcagctgatccctacttatttgtggcaaccatcgctgcatgttcatctcagaatgata 1557
Db 1640 TCAGCTGATTCCTCAC--CATCGTGGGACACATTTGCTGGAGCCCTCATCTTGTCTG 1697
QY 1558 attgattgatgttactagcaagaatcaaatbaacaaagcgaagatattgaagaagag 1617
Db 1698 ATTGCTTTATCTCTCT--GCGAGGTCAAGAACAAA--AGAAAGATGAGAGAGAG--AG 1754
QY 1618 aactgtatgcaagaagacttccaataatcctaaactgctgctgcaagcctcaccatct 1677
Db 1755 AGGCTGATTGAAGATGACTTCCATTAACCTAGACACTGAGG--CAGACCGGCTTCTCCAA--CT 1812
QY 1678 atgagcatatcaagcagctctcctcctcagtgtaagattacggctccaagaccgcta 1737
Db 1813 TAGGAGCGGACAAACAGCATCTTCCCAAAAGTCAGAAAGGGGTCCCGAGTCAGACCCCTA 1872
```

```
Search completed: March 30, 2002, 13:34:39
Job time: 23561 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 09:31:07 ; Search time 5438.31 Seconds
(without alignments)
636.253 Million cell updates/sec

Title: US-09-867-034-2

Perfect score: 322
Sequence: 1 gacagcaacaacccctga.....ctccaagtgtggaagact 322

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_estl:*
11: gb_estc:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_liv:*
17: em_gss_pla:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	321	99.7	441	10	AM364247 OV3-DT001
C 2	321	99.7	484	11	BF814154 RC3-CI004
C 3	321	99.7	489	11	BF513917 UT-H-BM1
C 4	321	99.7	619	10	AM161619 x146D06.x
C 5	321	99.7	623	11	BE867131 601442973
C 6	321	99.7	637	10	AM364300 OV3-DT001
C 7	321	99.7	666	11	BG778248 602666830
C 8	321	99.7	689	10	AM582256 OV4-SF021
C 9	321	99.7	689	10	BE748141 601571509
C 10	321	99.7	721	10	AM956284 EST368354
C 11	321	99.7	734	11	BE870718 601448570
C 12	321	99.7	742	11	BG823844 602729503

13	321	99.7	750	11	BG179160	BG179160 602330032
14	321	99.7	751	11	BG386151	BG386151 602455333
15	321	99.7	853	11	BG610804	BG610804 602612442
16	321	99.7	890	11	BG259927	BG259927 602371889
17	321	99.7	948	11	BE796856	BE796856 601588166
C 18	320	99.4	674	10	AI075324	AI075324 ov20a02.x
C 19	320	99.4	708	10	AI675865	AI675865 wB97h01.x
C 20	320	99.4	866	11	BG285246	BG285246 602409569
21	319.4	99.2	755	11	BF733142	BF733142 EST039 HU
C 22	319	99.1	595	10	AA573742	AA573742 HK07a09.S
C 23	319	99.1	744	10	AI815198	AI815198 wK72c03.x
C 24	319	99.1	782	10	AL543432	AL543432 AL543432
C 25	317.8	98.7	450	10	BE072059	BE072059 PM4-BT053
C 26	317.8	98.7	503	11	BF155131	BF155131 PM3-BT083
C 27	317.8	98.7	627	10	AM369331	AM369331 OV1-BN000
28	317	98.4	537	10	AA314225	AA314225 EST186091
C 29	316.8	98.4	643	10	AM151674	AM151674 x167e12.x
C 30	316.4	98.3	547	11	BE841679	BE841679 MR1-SN006
C 31	314.6	97.7	459	11	BF798311	BF798311 RC3-CI004
C 32	314.6	97.7	672	10	AI634717	AI634717 t209d10.x
C 33	312.2	97.0	566	10	AA316115	AA316115 EST187838
C 34	310.6	96.5	608	10	AI436796	AI436796 t182905.x
C 35	309.8	96.2	700	10	AI817063	AI817063 wJ76e07.x
C 36	309	96.0	740	11	BG501149	BG501149 602546389
C 37	308.8	95.9	659	10	AI380637	AI380637 t602d07.x
C 38	307.4	95.5	669	10	AI888294	AI888294 wN30e06.x
C 39	306.4	95.2	609	10	AM614754	AM614754 hN38a09.x
C 40	306	95.0	592	10	AA909945	AA909945 OM32908.S
C 41	305.4	94.8	551	10	AA307697	AA307697 EST178713
C 42	305	94.7	403	10	AM192785	AM192785 x151c07.x
C 43	303	94.1	631	10	AI080640	AI080640 oX54c07.x
C 44	303	94.1	682	10	AM194118	AM194118 xM12d06.x
C 45	302.6	94.0	475	11	BF744154	BF744154 PM3-BT083

ALIGNMENTS

RESULT 1
LOCUS AM364247 441 bp mRNA EST 04-FEB-2000
DEFINITION OV3-DT0012-081299-021-e06 DT0012 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM364247
VERSION AM364247.1 GI:6868897
KEYWORDS EST.
SOURCE human.
ORIGIN human.

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE HCGP http://www.ludwig.org.br/ORESTES.
JOURNAL The FAPESP/LICR Human Cancer Genome Project
COMMENT Unpublished (1999)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl-QV3&st-QV3-DT0012-081299-021-e06&t3=1999-12-08&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 440.

FEATURES

source location/Qualifiers
1..441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="DT0012"


```
/dev_stage="Adult"
/notes="Organ: denis_drash; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
```

BASE COUNT 114 a 90 c 96 g 141 t

ORIGIN

Query Match 99.7%; Score 321; DB 10; Length 441;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gaaagcaaaaccccttgatgattatcacccttgatgagtgccacacagtcagc 60
|||||
DB 424 GACACGACAAACCCTTGATGATTTATTCACCTTGAGAGAGCCACACAGTCAGC 365
|||||

OY 61 tttaagaaagtgcttgctgaataaagaatccagaatctgcaagacagttgtcct 120
|||||
DB 364 TTAAAGAAAGCTTTGCTGATAAATAAGAAATCCAGAAATTCGACAGAGTTTGCT 305
|||||

OY 121 cctcaatctggttatagaacaactgcaaacacctcttcctgtagtgccagatgtccc 180
|||||
DB 304 CCTCAATCTGGTTTATGAAACAACCTGACAAACACCTTCTCTGATGGCCAGTATGTC 245
|||||

OY 181 cagattatgcttgctgaaccatctcgcagattagagccgatatcaactggaagatctc 240
|||||
DB 244 CAGGATTATGTTGTTGACCCATCTGACAGATTAGCCGATATATCTGGAAGATTTC 185
|||||

OY 241 aaancgtctatagcttaagaacctgcagatacagctctgtgcttacaacatgaaga 300
|||||
DB 184 AATTCGCTCTATGCTTACGACACCTGACATACAGCTCTGTGCTTGACAACTGAAGA 125
|||||

OY 301 agctctcaagtgctgaagact 322
|||||

DB 124 AGCTCTCAAGTCTGCTGAAGACT 103
|||||

RESULT 2

BF814154 484 bp mRNA EST 12-JAN-2001
LOCUS
DEFINITION RC3-C10043-281100-025-b10 C10043 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF814154
VERSION BF814154.1 GI:12146229
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 484)
Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL.

(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC3&l2=RC3-C10043-
281100-025-b10&l3=2000-11-28&l4=1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 484.
Location/Qualifiers
1. 484
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="C10043"
/dev_stage="Adult"
/notes="Organ: colon_ins; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 148 a 114 c 95 g 127 t

ORIGIN

Query Match 99.7%; Score 321; DB 11; Length 484;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gacaagcaaaaccccttgatgattatcacccttgatgagtgccacacagtcagc 60
|||||
DB 97 GACACGACAAACCCTTGATGATTTATTCACCTTGAGAGAGTCCACACAGTCAGC 156
|||||

OY 61 tttaagaaagtgcttgctgaataaagaatccagaatctgcaagacagttgtcct 120
|||||
DB 157 TTAAAGAAAGCTTTGCTGATAAATAAGAAATCCAGAAATTCGACAGAGTTTGCT 216
|||||

OY 121 cctcaatctggttatagaacaactgcaaacacctcttcctgtagtgccagatgtccc 180
|||||
DB 217 CCTCAATCTGGTTTATGAAACAACCTGACAAACACCTTCTCTGATGGCCAGTATGTC 276
|||||

OY 181 cagattatgcttgctgaaccatctcgcagattagagccgatatcaactggaagatctc 240
|||||
DB 277 CAGGATTATGTTGTTGACCCATCTGACAGATTAGCCGATATATCTGGAAGATTTC 336
|||||

OY 241 aaancgtctatagcttaagaacctgcagatacagctctgtgcttacaacatgaaga 300
|||||
DB 337 AATTCGCTCTATGCTTACGACACCTGACATACAGCTCTGTGCTTGACAACTGAAGA 396
|||||

OY 301 agctctcaagtgctgaagact 322
|||||

DB 397 AGCTCTCAAGTCTGCTGAAGACT 418
|||||

RESULT 3

BF513917/c 489 bp mRNA EST 07-DEC-2000
LOCUS
DEFINITION UT-H-BW1-amq-f-02-0-UT.s1 NCI-CGAP_Sub7 Homo sapiens cDNA clone
IMAGE:3070970 3', mRNA sequence.
ACCESSION BF513917
VERSION BF513917.1 GI:11599096
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 489)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A


```
|||||
Db 317 CAGGATTATTTTGTGGACCACTCTGACAGTTAGAGCCGATATCATCTGGAAGATATTC 238
QY 241 aaangctctatgcttacgacaactcgagatagacgtctgtgcttgacacatgaagaa 300
|||||
Db 257 AATGCTGCTTATGCTTACGACACCTGACAGTACAGCTCTGTGCTTGACAACTGAAGAA 198
QY 301 agctctcaagttgctgaagact 322
|||||
Db 197 AGCTCTCAAGTGTGCTGAAGACT 176

RESULT 5
BE867131 BE867131 623 bp mRNA EST 20-OCT-2000
LOCUS 601442973F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846834 5'
DEFINITION mRNA sequence.
ACCESSION BE867131
VERSION BE867131.1 GI:10315907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://mgi.mci.nih.gov/
AUTHORS 1 (bases 1 to 623)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@rsf-mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9560 row: b column: 19
High quality sequence stop: 621.
Location/Qualifiers
1..623
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 Kb. Library constructed by Life
Technologies."
BASE COUNT 198 a 149 c 126 g 150 t
ORIGIN
Query Match 99.7%; Score 321; DB 11; Length 623;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaacaaccccttgatgatattatcatcacttgatgaagtcgacacagtaaacg 60
|||||
Db 194 GACACGACAAACCTTATGATATTTATTCATCTGATGAGTGTGCCACACAGTCAAGC 253
QY 61 tttaagaagaagtgttgcctggaataaagaatccagaatctgacagagcagttgtcct 120
|||||
Db 254 TTTAAAGAAAGTGTGCTGGAATAAAGAAATCCAGAAATTCGACAGACAGTTGTGCT 313
QY 121 cctcaatcgtgttatagaacaactgacaaacaccttctcctgagtgcagatgtcc 180
|||||
Db 314 CCTCAATCTGCTTATGAAGAACTGACAAACCTTCTCTCGATGAGCCAGATATGTCCC 373
QY 181 caggatattgttggaccacatctctgacagttagagccgatatatacagggaagatattc 240
|||||
```

```
Db 374 CAGGATTATTTTGTGGACCACTCTGACAGTTAGAGCCGATATCATCTGGAAGATATTC 433
QY 241 aaangctctatgcttacgacaactcgagatagacgtctgtgcttgacacatgaagaa 300
|||||
Db 434 AATGCTGCTTATGCTTACGACACCTGACAGTACAGCTCTGTGCTTGACAACTGAAGAA 493
QY 301 agctctcaagttgctgaagact 322
|||||
Db 494 AGCTCTCAAGTGTGCTGAAGACT 515

RESULT 6
AM364300/c AM364300 637 bp mRNA EST 04-FEB-2000
LOCUS QV3-DT0012-291299-051-g07 DT0012 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM364300
ACCESSION AM364300
VERSION AM364300.1 GI:6868950
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE HCGP http://www.ludwig.org.br/ORESTES
AUTHORS The FAPESP/LICR Human Cancer Genome Project
TITLE Unpublished (1999)
JOURNAL Contact: Simpson A.J.G.
COMMENT Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV3&t2=QV3-DT0012-
291299-051-g07&t3=1999-12-29&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 50.
High quality sequence stop: 609.
Location/Qualifiers
1..637
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lib="DT0012"
/dev_stage="Adult"
/notes="Organ: denis_drash; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 151 a 131 c 155 g 200 t
ORIGIN
Query Match 99.7%; Score 321; DB 10; Length 637;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaacaaccccttgatgatattatcatcacttgatgaagtcgacacagtaaacg 60
|||||
Db 467 GACACGACAAACCTTATGATATTTATTCATCTGATGAGTGTGCCACACAGTCAAGC 408
QY 61 tttaagaagaagtgttgcctggaataaagaatccagaatctgacagagcagttgtcct 120
|||||
Db 407 TTTAAAGAAAGTGTGCTGGAATAAAGAAATCCAGAAATTCGACAGACAGTTGTGCT 348
QY 121 cctcaatcgtgttatagaacaactgacaaacaccttctcctgagtgcagatgtcc 180
|||||
Db 347 CCTCAATCTGCTTATGAAGAACTGACAAACCTTCTCTCGATGAGCCAGATATGTCCC 288
|||||
```

QY 181 caggatattgttcttgaccatctctacaggttagagccatatactggaagatattc 240
 |||||
 Db 287 CAGGATTATGTTGTGTGACCATCTCTACAGATTAGACCCATATCACTGGAAAGTATTTC 228
 |||||
 QY 241 aaangctctctatcttgcagaaactgcagatacagctctgttcttgacaacatgaaga 300
 |||||
 Db 227 AAATGCTCTATGCTTATGCAAGCTGAGATACAGCTCTGTGCTTGACACATGAAGAA 168
 |||||
 QY 301 agctctcaagttgctgaagact 322
 |||||
 Db 167 AGCTCTCAAGTTGCTGAAGACT 146
 |||||

RESULT 7
 LOCUS BG778248 666 bp mRNA EST 15-MAY-2001
 DEFINITION 60266830F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4806679 5',
 mRNA sequence.
 ACCESSION BG778248
 VERSION BG778248.1 GI:14048565
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgaabs-remail.nih.gov
 Tissue Procurement: DCTD/DRP

CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 plate: LICM1656 row: 1 column: 08
 High quality sequence stop: 665.
 Location/Qualifiers

FEATURES
 source
 1..666

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4806679"
 /clone_id="NIH_MGC_60"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
 Site_1: SfiI (ggccgctggc); Site_2: SfiI (ggccattatggc
 5'; Double-stranded cDNA was prepared from cell line RNA.
 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCATTATGACC-3' and 3' adaptor
 sequence: 5'-ATCTAGAGCGCGCGCGCATG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
 library."
 BASE COUNT 205 a 165 c 140 g 156 t
 ORIGIN

Query Match 99.7%; Score 321; DB 11; Length 666;
 Best Local Similarity 99.7%; Pred. No. 8.7e-78;
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacagacaacaaacctgatgatattatcatcacttgatgaagccacacagtcacgc 60
 |||||
 Db 269 GACAAGCAACAACCTTGATGATTATTCATCACTTGATGATGCGCCACACAGTCAAGC 328
 |||||

QY 61 tttaagaagaatgttttcttgcaaaaataagaatccagaatttggcagagcagttgtcct 120
 |||||
 Db 329 TTTAAAGAAAGTGTGTTGCTGAAAAATAAGAAATCCAGAAATTTGGCAGACGATTGTCTT 388
 |||||
 QY 121 cctcaatcgtttatgaacaactgacaacacacctcttctctatgagccaglatgtccc 180
 |||||
 Db 389 CCTCAATCTGTTTATGAAACAACATGACAAACACCTTCTCCTGATGGCAGTATGTC 448
 |||||
 QY 181 caggatattgttcttgaccatctctcagaggttagagccgatatatactggaagatattc 240
 |||||
 Db 449 CAGGATTATGTTGTGTTGACCATCTCTGACAGTTTAGACCCATATCACTGGAAAGTATTTC 508
 |||||
 QY 241 aaangctctctatcttgcagaaactgcagatacagctctgttcttgacaacatgaaga 300
 |||||
 Db 509 AAATGCTCTATGCTTATGCAAGCTGCAATACAGCTCTGTGCTTGACACATGAAGAA 568
 |||||
 QY 301 agctctcaagttgctgaagact 322
 |||||
 Db 569 AGCTCTCAAGTTGCTGAAGACT 590
 |||||

RESULT 8
 LOCUS AM582256 689 bp mRNA EST 16-MAR-2000
 DEFINITION OVA-ST0212-120100-075-e10 ST0212 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AM582256
 VERSION AM582256.1 GI:7257305
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE HCEP http://www.ludwig.org.br/ORESTES.
 AUTHORS The FAPESP/LICR Human Cancer Genome Project
 TITLE Unpublished (1999)
 JOURNAL Contact: Simpson A.J.G.
 COMMENT Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV4<2=QV4-ST0212-
 120100-075-e10<3=2000-01-12<4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 687.
 Location/Qualifiers

FEATURES
 source
 1..689

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="ST0212"
 /dev_stage="Adult"
 /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. letters patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 212 a 172 c 143 g 162 t
 ORIGIN

Query Match 99.7%; Score 321; DB 10; Length 689;
 Best Local Similarity 99.7%; Pred. No. 8.7e-78;
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacagacaacaaacctgatgatattatcatcacttgatgaagccacacagtcacgc 60
 |||||

Db 282 GACAGCAACAAACCTTGATGATTATTCATCACTGGAGGAGTCCACACAGCTCAAGC 341
 QY 61 tttaagaagaagtgcttgctgaanaataagaanaatccagaanaatggcagagcagttgtcct 120
 Db 342 TTTTAAAGAAAGTGTTCCTGAAAAATAAGAAATCCAGAAATGGCAGAGCAGTTGTGCTT 401
 QY 121 ccttaactcgtgttatagaacaactgacaacaccttctcctgaagcagatgtgtcc 180
 Db 402 CCTCAATCTGGTTTATGTAACCAACAGCAACACCTTCTCCTGATGGCCAGTATGTGCC 461
 QY 181 cagattatcttgcttgaccatctcgaagctgaagccgatatcactgaaatattc 240
 Db 462 CAGATTATGTTGTGTGACCCATCTGACAGTTAGAGCCGATATCACTGGAATATTC 521
 QY 241 aaacgctctatgcttaagcaacctgacagatagcctgtgtgtgtgcttgaacaatgaaga 300
 Db 522 AAATCGTCTGATGCTTACGACACTGCAGATACAGCTGTGCTTGACAAACATGAAGAA 581
 QY 301 agctctcaagttgctgaagact 322
 Db 582 AGCTCTCAAGTTGCTGAAGACT 603

RESULT 9
 BE748141 689 bp mRNA EST 15-SEP-2000
 LOCUS 601571509F1 NIH_MGC_55 Homo sapiens cDNA IMAGE:3838347 5',
 DEFINITION mRNA sequence.
 BE748141
 ACCESSION BE748141.1 GI:10162133
 VERSION
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 689)
 AUTHORS NIH-MGC. <http://mgc.nci.nih.gov/>.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLCM25 row: a column: 04
 High quality sequence stop: 643.
 Location/Qualifiers
 1. 689
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3838347"
 /clone_lib="NIH_MGC_55"
 /tissue_type="from acute myelogenous leukemia"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
 Site_1: SfiI (ggcgctcgagcc); Site_2: SfiI (ggcattagagcc
); Double-stranded cDNA was prepared from cell line RNA.
 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGCGCATATGCGC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGCGCGCGCGCATG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.65 kb (range 0.9-4.0 kb). 14/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."

BASE COUNT 209 a 170 c 148 g 162 t

Query Match 99.7%; Score 321; DB 10; Length 689;
 Best Local Similarity 99.7%; Pred. No. 8.7e-78;
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaaaacccctgattgattatctacactgtagtgagtgccacacagtcagc 60
 Db 251 GACAGCAACAAACCTTGATGATTATTCATCACTGGAGGAGTCCACACAGCTCAAGC 310
 QY 61 tttaagaagaagtgcttgctgaanaataagaanaatccagaanaatggcagagcagttgtcct 120
 Db 311 TTTTAAAGAAAGTGTTCCTGAAAAATAAGAAATCCAGAAATGGCAGAGCAGTTGTGCTT 370
 QY 121 ccttaactcgtgttatagaacaactgacaacaccttctcctgaagcagatgtgtcc 180
 Db 371 CCTCAATCTGGTTTATGTAACCAACAGCAACACCTTCTCCTGATGGCCAGTATGTGCC 430
 QY 181 cagattatcttgcttgaccatctcgaagctgaagccgatatcactgaaatattc 240
 Db 431 CAGATTATGTTGTGTGACCCATCTGACAGTTAGAGCCGATATCACTGGAATATTC 490
 QY 241 aaacgctctatgcttaagcaacctgacagatagcctgtgtgtgtgcttgaacaatgaaga 300
 Db 491 AAATCGTCTGATGCTTACGACACTGCAGATACAGCTGTGCTTGACAAACATGAAGAA 550
 QY 301 agctctcaagttgctgaagact 322
 Db 551 AGCTCTCAAGTTGCTGAAGACT 572

RESULT 10
 AW956284 721 bp mRNA EST 01-JUN-2000
 LOCUS AW956284
 DEFINITION EST368354 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW956284
 VERSION AW956284.1 GI:8145967
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 721)
 AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt
 I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
 Quackenbush, J.
 JOURNAL Assessment of gene expression patterns in a model of colon tumor
 metastasis using a 19,200 element cDNA microarray
 COMMENT Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@igrr.org
 Plate: 90
 Seq primer: Reverse.
 Location/Qualifiers
 1. 721
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MAGE resequences, MAGD"
 /note="Vector: pBluescriptSkm" 1 others

BASE COUNT 224 a 173 c 148 g 175 t

ORIGIN
 Query Match 99.7%; Score 321; DB 10; Length 721;
 Best Local Similarity 99.7%; Pred. No. 8.7e-78;
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaaaacccctgattgattatctacactgtagtgagtgccacacagtcagc 60
 Db 260 GACAGCAACAAACCTTGATGATTATTCATCACTGGAGGAGTCCACACAGCTCAAGC 319

Qy 61 tttaagaagaatgttgcgtgaataaataaagaataatccagaataatggcagagcagttgtcct 120
|||||
Db 320 TTTAAAGAAAGTGTGCTGAAAAATTAAGAAATTCAGAAATTTGGCAGAGCAGTTGTCT 379
|||||
Qy 121 cctcaatcgtgttatgtaaaacactgcaaacacttctcctgtagtgcagtagtccc 180
|||||
Db 380 CCTCAATCTGGTTTATGAAACACGACAAACACCTTCTCCTGATGGCCAGTATGTCC 439
|||||
Qy 181 caggatitgcttgcgtcccatctctgacagtiagagccgatatcactggaagaatctc 240
|||||
Db 440 CAGGATTTATGTTGTTGACCCATCTCTGACAGTTAGACCCGATATCCTGGAGATATTC 499
|||||
Qy 241 aaancgtctctatgcttcaagaacctgcagatacagctctgttgccttgacaacatgaaga 300
|||||
Db 500 AAATCGTCTGATGTTAGCAACCTGCAGATACACCTCTGTTGCTTGACACATGAAGA 559
|||||
Qy 301 agctctcaagtgtcgtgaagact 322
|||||
Db 560 AGCTCTCAAGTTGCTGAAGACT 581
|||||

RESULT 11
BE870718 734 bp mRNA EST 20-OCT-2000
LOCUS 601448570F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852448 5',
DEFINITION mRNA sequence.
ACCESSION BE870718
VERSION BE870718.1 GI:10319507
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 734)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
plate: LLM9574 row: 1 column: 17
High quality sequence stop: 723.
Location/Qualifiers
source 1..734
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3852448"
/clone_id="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site:1; Not:
Site-2; Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 Kb. Library constructed by Life
Technologies."

BASE COUNT 231 a 163 c 156 g 184 t
ORIGIN
Query Match 99.7%; Score 321; DB 11; Length 734;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gacaagcaacaacccctgatgatcatcatcactgtagtgcacacacagtcagc 60
|||||
Db 232 GACAGCAACAACCCCTTGATGATATTCATCATTGGATGGATGCCACACAGTCAGC 291
|||||
Qy 61 tttaagaagaatgttgcgtgaataaataaagaataatccagaataatggcagagcagttgtcct 120
|||||

Db 292 TTTAAAGAAAGTGTGCTGAAAAATTAAGAAATTCAGAAATTTGGCAGAGCAGTTGTCT 351
|||||
Qy 121 cctcaatcgtgttatgtaaaacactgcaaacacttctcctgtagtgcagtagtccc 180
|||||
Db 352 CCTCAATCTGGTTTATGAAACACGACAAACACCTTCTCCTGATGGCCAGTATGTCC 411
|||||
Qy 181 caggatitgcttgcgtcccatctctgacagtiagagccgatatcactggaagaatctc 240
|||||
Db 412 CAGGATTTATGTTGTTGACCCATCTCTGACAGTTAGACCCGATATCCTGGAGATATTC 471
|||||
Qy 241 aaancgtctctatgcttcaagaacctgcagatacagctctgttgccttgacaacatgaaga 300
|||||
Db 472 AAACGCTCTGATGTTAGCAACCTGCAGATACACCTCTGTTGCTTGACACATGAAGA 531
|||||
Qy 301 agctctcaagtgtcgtgaagact 322
|||||
Db 532 AGCTCTCAAGTTGCTGAAGACT 553
|||||

RESULT 12
BG823844 742 bp mRNA EST 22-MAY-2001
LOCUS 602729503P1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4878336 5',
DEFINITION mRNA sequence.
ACCESSION BG823844
VERSION BG823844.1 GI:14171431
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 742)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
plate: LUCM1764 row: f column: 01
High quality sequence stop: 702.
Location/Qualifiers
source 1..742
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4878336"
/clone_id="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site:1; XhoI; Site-2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 232 a 178 c 151 g 181 t
ORIGIN
Query Match 99.7%; Score 321; DB 11; Length 742;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gacaagcaacaacccctgatgatcatcatcactgtagtgcacacacagtcagc 60
|||||

Db 239 GACACGACAAACCCCTTGATGATTATTCATCACTGGATGAGGCCACACAGTCAAGC 298
Qy 61 tttaagaagaagtgttctgtaaaaaataagaataatccagaatctgacagcagttgtcct 120
Db 299 TTTAAAGAAAGTGTTCCTGTAATAAAGAAATCCAGAAATGGCAGAGCAGTTGTGCT 358
Qy 121 ccttaactcgtgttataagaacaactgacaacaccccttctcctgtagtgccagtagtccc 180
Db 359 CCTCAATCTGGTTTATGAAACAACTGACAAACACCTTCTCCTGATGGCCAGATATGTC 418
Qy 181 cagattatgttctgtgacccatctcgacagttagagccgatatcaactggaagatattc 240
Db 419 CAGATTATGTTTGTTCACCATCTTGACAGTTAGAGCCGATATCACTGGAAGATATTC 478
Qy 241 aaanagctctatgtctacgaacctgcagatacagctctgtgtctgttgcatacaacagaa 300
Db 479 AAATGCTCTATGCTTACGAACTGCAGATACAGCTCTGTGCTTGACAAATGAAAGAA 538
Qy 301 agctctcaagtgtcgaagact 322
Db 539 AGCTCTCAAGTGTCTGAAGACT 560

RESULT 13

BG179160 750 bp mRNA EST 06-FEB-2001
LOCUS 602330032F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4431583 5',
DEFINITION mRNA sequence.
BGI79160

ACCESSION BG179160 GI:12685943
VERSION BG179160.1
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 750)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DRP

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLAM10186 row: O column: 08
High quality sequence stop: 732.
Location/Qualifiers

FEATURES

source 1..750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4431583"
/clone_lib="NIH_MGC_91"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."
BASE COUNT 234 a 174 c 158 g 183 t 1 others
ORIGIN

Query Match 99.7%; Score 321; DB 11; Length 750;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 gacacgaacaaaccccttgatgattatcattcactgtagtgagtcacacagtcgaagc 60
|||||

Db 263 GACACGACAAACCCCTTGATGATTATTCATCACTGGATGAGGCCACACAGTCAAGC 322
Qy 61 tttaagaagaagtgttctgtaaaaaataagaataatccagaatctgacagcagttgtcct 120
Db 323 TTTAAAGAAAGTGTTCCTGTAATAAAGAAATCCAGAAATGGCAGAGCAGTTGTGCT 382
Qy 121 ccttaactcgtgttataagaacaactgacaacaccccttctcctgtagtgccagtagtccc 180
Db 383 CCTCAATCTGGTTTATGAAACAACTGACAAACACCTTCTCCTGATGGCCAGATATGTC 442
Qy 181 cagattatgttctgtgacccatctcgacagttagagccgatatcaactggaagatattc 240
Db 443 CAGATTATGTTTGTTCACCATCTTGACAGTTAGAGCCGATATCACTGGAAGATATTC 502
Qy 241 aaanagctctatgtctacgaacctgcagatacagctctgtgtctgttgcatacaacagaa 300
Db 503 AAATGCTCTATGCTTACGAACTGCAGATACAGCTCTGTGCTTGACAAATGAAAGAA 562
Qy 301 agctctcaagtgtcgaagact 322
Db 563 AGCTCTCAAGTGTCTGAAGACT 584

RESULT 14

BG386151 751 bp mRNA EST 12-MAR-2001
LOCUS 60245333F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:458336 5',
DEFINITION mRNA sequence.
BG386151

ACCESSION BG386151 GI:13279597
VERSION BG386151.1
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 751)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLCM1307 row: b column: 09
High quality sequence stop: 672.
Location/Qualifiers

FEATURES

source 1..751
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:458336"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(C). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 230 a 181 c 153 g 187 t
ORIGIN

Query Match 99.7%; Score 321; DB 11; Length 751;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Accession	Version	Keywords	Source	Organism	Reference Authors	Title	Journal	Comment
Db	232	GACAGCAACAAACACCTTGATGATGATTTATTCATCCTCTGGATAGTGGCCACACATCAAGC	291					
Qy	61	tttaagaagaagtggttctgctaataaagaagaatccagaatctgacagacagttgacct	120					
Db	292	TTTAAAGAAAGCTGTTTGGCTGAATAATTAAGAAATCCAGAAATTTGGCAGACAGCTTTTGCTCT	351					
Qy	121	ccctaatcttggtttatagaaacaactgacaaacaccttctctctgctgagtcagatgctcc	180					
Db	352	CCTCAATGTGGTTTATGTAACAAACATGACAAACACCTTTCTCTGATGGCCAGTATGTC	411					
Qy	181	cagagatatgtttgttgaccccatctctgacagcttagagccgatatcaactggaagatctc	240					
Db	412	CAGGATTTATGTTTGTGGACCCATCTCTGACAGTTAGAGCCCATATTCACCTGGAAACATATTC	471					
Qy	241	aaanagctctcatgcttgcttcggaacctgcagagctgaagagctctgttcttgcttacaacatgaaga	300					
Db	472	AAACCGTCTCTATGCTTACGAACCTCGCAGATACAGCTCTGTGCTTGAACAACATGAAGAA	531					
Qy	301	agctctcaagctgctcgaagact	322					
Db	532	AGCTCTCAAGTTGCTGAGAGACT	553					
RESULT	15							
LOCUS	BG610804	853 bp	mRNA	EST	18-APR-2001			
DEFINITION	60261244ZP1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4737755	5'						
ACCESSION	BG610804							
VERSION	BG610804.1	GI:13662175						
KEYWORDS	EST.							
SOURCE	human.							
ORGANISM	Homo sapiens							
REFERENCE	Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.							
TITLE	1 (bases 1 to 853)							
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/							
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)							
	Unpublished (1999)							
	Contact: Robert Strausberg, Ph.D.							
	Email: cgapbs-remail.nih.gov							
	Tissue Procurement: DCM/DTP							
	CDNA Library Preparation: CLONTECH Laboratories, Inc.							
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)							
	DNA Sequencing by: Incyte Genomics, Inc.							
	Clone distribution: MGC clone distribution information can be							
	found through the I.M.A.G.E. Consortium/LNL at:							
	http://image.lnl.gov							
	Plate: LNCM1606	row: d	column: 12					
	High quality sequence stop: 812.							

```

FEATURES
    source
        Location/Qualifiers
            1..853
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_image="473755"
                /clone_lib="NIH_MGC_60"
                /tissue_type="adenocarcinoma"
                /lab_host="DH10B (T1 phage-resistant)"
                /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                Site_1: SfiI (ggcgccctcgccg); Site_2: SfiI (ggccatcattggccg);
                Double-stranded cDNA was prepared from cell line RNA.
                5' and 3' adaptors were used in cloning as follows: 5'
                adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                sequence: 5'-ATTCTAGAGAGCGGAGCGCGGCACATG-dT(30)NN-3'
                (where B = A, C, or G and N = A, C, G, or T). Average
                insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
                contained inserts by PCR. This library was enriched for
                full-length clones and was constructed by Clontech
                Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                library."

```

BASE COUNT	273 a	193 c	174 g	213 t
ORIGIN				
Query Match		99.7%;	Score 321;	DB 11;
Best Local Similarity		99.7%;	Pred. No. 8.7e-78;	
Matches 321; Conservative		0;	Mismatches 1;	Indels 0; Gaps 0;
Db	1	gacagcagacaaaccccttgatgatattatcattacttcttgatggtgccacaaagtcgaacg	60	
Db	253	GCAAGCAACAACCCCTTGATGATTTATCTACTGTGATGAGTGCCCAACAGTCAAC	312	
Qy	61	tttaaagaagtgcttctgcgtaaaataaagaatccagaatgagcagcttgctct	120	
Db	313	TTTAAAGAAAGTGTTGCGTAATAAAGAAATCCGAATTTGGCAGACACTTTGTCT	372	
Qy	121	cctcaatctggtttatagaaacaactgcacaacaccttctcctgatgacgattgtccc	180	
Db	373	CCTCAATCTGGTTTATGAAACAACGTACAACACCTTTCTCTGATGGCCAGTATGTCC	432	
Qy	181	caggaattatgcttctgacccatcctctgcagctttagagccgatactcaactggaagatattc	240	
Db	433	CAGCAATTAGTTGTTGTGACCACTCTCTGCACTTAGAGCCGATATCACCGAAGATATTTC	492	
Qy	241	aaanagctctcatgtccttgcagaaactgcagatatacagctctgttgccttgcagacaataaaga	300	
Db	493	AAATGCTCTCTATGCTTAGCAACCTGCACATACAGCTCTGTGCTTGACACATGAGAA	552	
Qy	301	agctctcaagttctgcgaagact	322	
Db	553	AGCTCTCAAGTTGCTGAAAGACT	574	

Search completed: March 30, 2002, 09:31:11
Job time: 13928 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 07:01:58 ; Search time 8456.85 seconds

(without alignments)
1777.131 Million cell updates/sec

Title: US-09-867-034-1

Sequence: 1 tttttttttgcgtgttt.....cacgcgtggcctccggagcgc 911

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	607	66.6	57000	9	AB038162 Homo sapi
2	607	66.6	157739	9	AP001623 Homo sapi
3	607	66.6	181511	9	AC015555 Homo sapi
4	607	66.6	340000	9	AP001555 Homo sapi
5	516.8	56.7	834	6	AX017268 Sequence
6	516.8	56.7	834	6	AX018086 Sequence
7	403	44.2	480	6	AX193486 Sequence
8	403	44.2	480	6	HUMPIBX
9	305.4	33.5	570	6	195749
10	264.4	29.0	398	6	HUMREFAC
11	236.2	25.9	250	6	AX106398
12	236.2	25.9	250	6	AX140689
13	220.6	24.2	248	6	AX106296
14	220.6	24.2	248	6	AX140587
15	195	21.4	320	9	HSHTF3
16	191	21.0	730	9	HSHTF2
17	191	21.0	68531	2	AC055868
18	182	20.0	182	11	G06131
19	182	20.0	182	11	G34583
20	171	18.8	277	6	AX062739
21	134.4	14.8	206	6	AX062644
22	108.8	11.9	450	10	MUSMITF
23	108.8	11.9	472	10	BC011042
24	108.8	11.9	7035	10	MMU271004
25	105.2	11.5	424	10	RNU48825
26	103	11.3	184	10	AF012534
27	101.2	11.1	5116	10	MMU46858
28	100.6	11.0	264	10	RNIF
29	100.6	11.0	264	10	S49317
30	100.4	11.0	431	10	RATREFOIL
31	94.8	10.4	1128	9	HSHTF1
32	92.4	10.1	144	10	MMTPEX2
33	64	7.0	6779	10	MMU271003
34	63.6	7.0	574	4	SESPAA
35	61.8	6.8	546	5	XELXP2A
36	61.2	6.7	1471	5	XIAPPG
37	60.2	6.6	447	10	MMPS2MPDP
38	60	6.6	546	10	MMSPAA
39	60	6.6	8241	6	AX069291
40	59.6	6.5	92580	2	RN141N5
41	59.2	6.5	5505	6	AX069290
42	58.8	6.5	466	10	RATPS2P
43	58.8	6.5	5142	6	AX069289
44	58.8	6.5	7048	10	MMU271002
45	57.6	6.3	4405	10	MMU78770

ALIGNMENTS

RESULT 1
AB038162/c
LOCUS
DEFINITION Homo sapiens TTF gene cluster for trefoil factor, complete cds.
ACCESSION AB038162
VERSION AB038162.1 GI:10280533
KEYWORDS trefoil factor.
SOURCE Homo sapiens pre-pro-B cell cell_line:FLB14-14 DNA, clone_lib:Keio
BAC library clone:K8169B4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)

REFERENCE

AUTHORS

Berry,A., Scott,H.S., Kudoh,J., Tailor,I., Korostishevsky,M.,
Wattenhofer,M., Guipponi,M., Barras,C., Rossier,C., Shinyu,K.,
Wang,J., Kawasaki,K., Asakawa,S., Minoshima,S., Shimizu,N.,
Antonarakis,S. and Bonne-Tamir,B.
Refined localization of autosomal recessive nonsyndromic deafness
DFNB10 locus using 34 novel microsatellite markers, genomic
structure, and exclusion of six known genes in the region

TITLE

JOURNAL Genomics 68 (1), 22-29 (2000)
 MEDLINE 20408883
 REFERENCE 2 (bases 1 to 57000)
 AUTHORS Shimizu, N., Kudo, J. and Shibuya, K.
 TITLE Direct Submission
 JOURNAL Submitted (10-FEB-2000) to the DDBJ/EMBL/Genbank databases.
 Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
 Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
 (E-mail:nshimizu@med.keio.ac.jp, Tel:81-3-3351-2370,
 Fax:81-3-3351-2370)
 location/Qualifiers
 1. 57000

FEATURES
 source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="FLEB14-14"
 /cell_type="pre-pro-B cell"
 /chromosome="21"
 /clone_1lb="KB169B4"
 /map="21q22.3"
 /map="Keio BAC library"
 mRNA join(957..1081,4085..4237,4997..5208)
 gene /gene="TFP1"
 957..5208
 CDS join(997..1081,4085..4237,4997..5013)
 /gene="TFP1"
 /codon_start=1
 /product="trefol1 factor 1"
 /protein_id="BABI3729.1"
 /db_xref="GI:10280534"
 /translation="MATMKVICALVIVSMALGTLAEQETETCTVAPRERONCGFP
 GYTPSOCANKGCCFDDTVKGVPCFYPNTIDVPEECF"
 mRNA join(16511..16613,17462..17611,19860..20006,20946..21132)
 gene /gene="TFP2"
 16511..21132
 CDS join(16535..16613,17462..17611,19860..20006,20946..20959)
 /gene="TFP2"
 /codon_start=1
 /product="trefol1 factor 2"
 /protein_id="BABI3730.1"
 /db_xref="GI:10280535"
 /translation="MGRDADLIALLVIGLALAGSEKSPQCSRLSPHNNTNGF
 PGITSDQCPDNGCCFSSVTGVPCFHPKQSDQCVMEVSRNRMCIGPISPECA
 SRKCCSFNLEFVPCFPRSVEDCH"
 mRNA join(52117..52198,53860..54006,55222..55412)
 gene /gene="TFP3"
 52117..55412
 CDS join(52138..52198,53860..54006,55222..55235)
 /gene="TFP3"
 /codon_start=1
 /product="trefol1 factor 3"
 /protein_id="BABI3731.1"
 /db_xref="GI:10280536"
 /translation="MGGLVIALSSSSAEERYVLSANQCAVPAKDRVDCGYPVHTKE
 CNRGCFCFSRLPGVPCFPLDEACTP"
 BASE COUNT 13757 a 13698 c 13997 g 15548 t
 ORIGIN

Query Match 66.6%; Score 607; DB 9; Length 57000;
 Best Local Similarity 100.0%; Pred. No. 1.1e-162;
 Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tgcctgttctcataatgttactgtacaaagaaacccaggaatagtagcaagta 71
 |||||||
 DB 55824 TGCCTGTTTCTCTATATGTTTACTGTACAAAGAAACCAACCCAGGATAGTACAAATA 55765
 |||||||

OY 72 ttgaacagtagcagagtggtctgtgaataaagagcaacttggaaagacagtttattgg 131
 |||||||
 DB 55764 TTGAACAGTAGCAGAGTGGTGTGAATAAAGAGCAACCTTTGGAACACAGTTTATTGG 55705
 |||||||

OY 132 ctgtgtcttcacccaagaagaagcttgatgtttttgaaactcttactgaatgtattt 191
 |||||||
 DB 55704 CTGTGCTCTTTCACCAAGAAAGACTGTGATTTTGAATAAATCTTACTTAATAATGATT 55645
 |||||||

OY 192 ttctgtcttcocaggaagcagcaacttacagtgcttcttaggtcttcgttgacgtgagt 251
 |||||||
 DB 55644 TTCTGTCTTCCGAGAGACGGGACACTTACAGTGTCTTCTAGGCTTCTGTACGTGGT 55585
 |||||||

OY 252 gccagcttgatccaataatccttgatgacatgacgtcctccttaggagcttcttcgt 311
 |||||||
 DB 55584 GCCAGTCTGATTCANAATATCTCTTGATGCACAGCTCCTTAGGAGACTTTTCTCTG 55525
 |||||||

OY 312 ccccttagagcctggagagactctcccttgacacctcccgccctctccagcgagca 371
 |||||||
 DB 55524 CCTTGGAGCCTGGGAGACTCTCCCTGACACCTCTCCCTCTCCACGAGCGAGCA 55465
 |||||||

OY 372 gaataaagcacaaacctcagaagctcagagcagaaagacttccctgggtgagcatg 431
 |||||||
 DB 55464 GAATTAAGCACAACTCAGAAAGTCTCAGGACGAGAAACTGTCTCGGGTGGAGCATG 55405
 |||||||

OY 432 ggaacttattcgtttaagacatcagagctcagatatgaacttccagagaagcgcttgc 491
 |||||||
 DB 55404 GGAACCTTATTCTGTTAAGACATCAGCTCCAGATATGAACCTTCAGACAGAGCGTTGGC 55345
 |||||||

OY 492 gggagcagaagggagcagaagaagctgagatgacagtgcttgcagcaatcacagcgagca 551
 |||||||
 DB 55344 GGGAGCAAGAGGACAGAAAGCTGAGATGAACGTGCTGGCAGCAATCCACCGGCA 55285
 |||||||

OY 552 aggggtctcgaagcctcgatcccccggcggggagcgagctggaggtgctcagaaggtgc 611
 |||||||
 DB 55284 AGGGTCTCCGAGAGCTTCGATCCCGCGCGGGGAGCTGGAGAGTCCCTCAGAAAGTGC 55225
 |||||||

OY 612 attctgc 618
 |||||||
 DB 55224 ATTCTGC 55218
 |||||||

RESULT 2
 AP001623 DNA PRI 03-JUN-2000
 LOCUS Homo sapiens genomic DNA, chromosome 21, clone:KB169B4, MX1-D21s171
 DEFINITION region, complete sequence.
 ACCESSION AP001623
 VERSION AP001623.1 GI:7670577
 KEYWORDS HTG.
 SOURCE Homo sapiens pre-pro-B cell cell_line:FLEB14-14 DNA, clone_1lb:Keio
 BAC library clone:KB169B4.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 157739)
 AUTHORS Shimizu, N., Kudo, J. and Shibuya, K.
 TITLE Homo sapiens genomic DNA, chromosome 21, clone:KB169B4, MX1-D21s171
 JOURNAL Published Only in DataBase (2000) In press
 2 (bases 1 to 157739)
 DIRECT SUBMISSION
 TITLE Submitted (04-APR-2000) to the DDBJ/EMBL/Genbank databases.
 JOURNAL Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
 Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
 (E-mail:nshimizu@med.keio.ac.jp, Tel:81-3-3351-2370,
 Fax:81-3-3351-2370)
 location/Qualifiers
 1. 157739
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="FLEB14-14"
 /cell_type="pre-pro-B cell"
 /chromosome="21"
 /clone_1lb="KB169B4"
 /clone_1lb="Keio BAC library"
 BASE COUNT 41302 a 39955 c 38798 g 37684 t

ORIGIN

Query Match

66.6%: Score 607; DB 9; Length 157739;

Best Local Similarity 100.0%: Pred. No. 1.2e-162; Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tgcctgttgcataatgttactgtacaaagaacaaacccagatagtaagta 71
 |||||||
 Db 42041 TGCGTGTGTTCAATAGTTACTGTACAAAGAAACAAACCCAGGATAGTCAAGTA 42100
 |||||||
 QY 72 ttgaacagtagcgaagtggtgtgaataaagaacacatttgaagaacagttatttg 131
 |||||||
 Db 42101 TTGAACAGTAGCAGAGAGGCTGTGAATAAAGAACCCCTTGGAAAGACAGCTTTATTTGG 42160
 |||||||
 QY 132 ctgctgtcttcacacaagaagaacttggatttttgaanaacttactactgaatgattt 191
 |||||||
 Db 42161 CTGCTGTCTTCACCAAAAGACTTGATTTTGAATACTTACTGAATGTATTT 42220
 |||||||
 QY 192 ttctgtcttcacagaagcagcaacttacttctcctgtgaagtggt 251
 |||||||
 Db 42221 TTTCTGCTTCCGAGAGGCGCACTTACAGTGTCTTCTAGGCTTCTGTGAGCTGGGT 42280
 |||||||
 QY 252 gccagctgtgaatcaaatatccctgtcagctgcagctcctcttaaggagtcctttcctg 311
 |||||||
 Db 42281 GCCAGTCTGATCAAAATATCTTGATGATGCACTGCACCTCTTAGGAGTCTTTCTGTG 42340
 |||||||
 QY 312 cccctgagggcctggcagcagctccctcctgacacccctccgcctctcccaagcagcagca 371
 |||||||
 Db 42341 CCCCTTGAGGCGCTGGGCGAGACTCTCCCTCGAACCCCTCCGCTCCACAGCAGCAGCA 42400
 |||||||
 QY 372 gaataaagcacaaccccaagaagcttcagcagcagaagaactgtccctggttgagacatg 431
 |||||||
 Db 42401 GAATTAAGCACAACCTCAGAAAGTCTCAGGCAGCAAGAACTGTCTGGGTTGAGCATG 42460
 |||||||
 QY 432 ggaaccttatctgttaagacatcagcgtccagatatgaacttcaagcagaagcagctggc 491
 |||||||
 Db 42461 GGACCTTTATTCGTTAAGACATCAGGCTCCAGATATGAACTTTCAGCAGAAAGCCGTTGCC 42520
 |||||||
 QY 492 gggagcagaagagacagaagaagctgagatgaacatgtcctgtgcagcaatcaagccggagca 551
 |||||||
 Db 42521 GGGAGCAAAAGGACAGAAAGCTGATGATGACAGTGGCTGGCAGCAATCAACAGCCGGGCA 42580
 |||||||
 QY 552 agggatgtcgcagcagctgcaccccgccggcgggagcagctgagagtgctcctaagaagtgcc 611
 |||||||
 Db 42581 AGGGTGCTCCAGCTCGCATCCCGGGCGGGGCGACACTGAGAGTGCTCAGAAAGTGTC 42640
 |||||||
 QY 612 attctgc 618
 |||||||
 Db 42641 ATTCTGC 42647

RESULT 3
 AC015555/c 181511 bp DNA PRI 14-DEC-2000
 LOCUS Homo sapiens chromosome 21 clone RP11-113F1, complete sequence.
 DEFINITION AC015555
 ACCESSION AC015555.13 GI:11602942
 VERSION
 KEYWORDS HMG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 181511)
 AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
 Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
 Mao,J., Lam,B., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,
 Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmy,J.,
 Yu,S. and Davis,R.W.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 181511)
 Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,
 Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,

TITLE Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
 Yu,S. and Davis,R.W.
 JOURNAL Direct Submission
 Submitted (17-NOV-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 REFERENCE 3 (bases 1 to 181511)
 AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
 Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
 Lam,B., Mao,J., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,
 Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmy,J.,
 Yu,S. and Davis,R.W.
 JOURNAL Direct Submission
 Submitted (14-DEC-2000) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 COMMENT On Dec 8, 2000 this sequence version replaced gi:11386278.
 ----- Genome Center
 Center: Stanford DNA Sequencing and Technology Development
 Center
 Center code: SDSFDC
 Web site: <http://sequence-www.stanford.edu/group/human/>
 Contact: hum.info@sequence.stanford.edu
 ----- Project Information
 Center project name: 694
 Center clone name: RP11-113F1

----- Summary
 This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

Bases 1-126158 of this BAC clone overlap with the complement of
 bases 1-126554 of KB169B4 (AP001623) and bases 105422-181511 of
 this BAC clone overlap with the complement of bases 73187-149850 of
 KB1430A10 (AP001622).
 Location/Qualifiers
 1. 181511
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /clone="RP11-113F1"
 /clone_1lb="RP11 human BAC library 11"
 102209
 /note="Single-stranded, single chemistry with consensus
 quality below 30."
 102988..102990
 /note="Single-stranded, single chemistry with consensus
 below 30."
 125544
 /note="Single-stranded, single chemistry with consensus
 quality below 30."
 125566
 /note="Single-stranded, single chemistry with consensus
 quality below 30."
 166246..166247
 /note="Single-stranded, single chemistry with consensus
 quality below 30."
 misc_feature
 BASE COUNT 43836 a 46156 c 45935 g 4584 t
 ORIGIN

Query Match 66.6%: Score 607; DB 9; Length 181511;
 Best Local Similarity 100.0%: Pred. No. 1.2e-162;
 Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tgcctgttgcataatgttactgtacaaagaacaaacccagatagtaagta 71
 |||||||
 Db 84525 TGCGTGTGTTCAATAGTTACTGTACAAAGAAACAAACCCAGGATAGTCAAGTA 84466

```

OY      72  ttgacacgtacgagagtggtgtgaataaaggaccatttgaagacagtttattg 131
         |||||||
Db      84465  ttgaacagtagcagagtggtgtgaataaaggaccatttgaagacagtttattg 84406
OY      132  ctgtgtcttcttcccaagaagaagctgtgtatttltgaagaacttactgaatgattt 191
         |||||||
Db      84405  ctgtgtcttcttcccaagaagaagctgtgtatttltgaagaacttactgaatgattt 84346
OY      192  ttctgtcttcttcccaagaagaagctgtgtatttltgaagaacttactgaatgattt 251
         |||||||
Db      84345  ttctgtgtcttcttcccaagaagaagctgtgtatttltgaagaacttactgaatgattt 84286
OY      252  gccagtcgtacatcaaaatccttgcatacgtacgtacgtcctccttgaagagcttctctg 311
         |||||||
Db      84285  gccagtcgtacatcaaaatccttgcatacgtacgtacgtcctccttgaagagcttctctg 84226
OY      312  ccccttgagcctgggagagctcttcccttgacacctccgccccttcccaagaagcagca 371
         |||||||
Db      84225  ccccttgagcctgggagagctcttcccttgacacctccgccccttcccaagaagcagca 84166
OY      372  gaataaagacacaacctcagaagaagctcagcagcagaagctcctcaggtgagcagtg 431
         |||||||
Db      84165  gaataaagacacaacctcagaagaagctcagcagcagaagctcctcaggtgagcagtg 84106
OY      432  ggaccttattcgttaagacatcaggtcctcagatatgaacttcaagagaagcagcttgc 491
         |||||||
Db      84105  ggaccttattcgttaagacatcaggtcctcagatatgaacttcaagagaagcagcttgc 84046
OY      492  gggagcgaagagcagaagaagctgagatgaacagtgcttgccagcaatacagcagcgga 551
         |||||||
Db      84045  gggagcgaagagcagaagaagctgagatgaacagtgcttgccagcaatacagcagcgga 83986
OY      552  aggggtgctcgcagcctgcatacccccgcggcgagcgagctggaagtgctcctcagaagtg 611
         |||||||
Db      83985  aggggtgctcgcagcctgcatacccccgcggcgagcgagctggaagtgctcctcagaagtg 83926
OY      612  attctcgc 618
         |||||||
Db      83925  attctcgc 83919

RESULT  4
AP001746 340000 bp DNA PRI 30-MAY-2000
LOCUS Homo sapiens genomic DNA, chromosome 21q, section 90/105.
ACCESSION AP001746 AL163291 BA000005
VERSION AP001746.1 GI:7768741
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
          1 (sites)
REFERENCE Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
AUTHORS Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
          Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
          Polley,A., Menzel,U., Delabar,U., Kumpf,K., Lehmann,R.,
          Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
          Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
          Asakawa,S., Shintani,A., Sasaki,T., Nagamine,N., Mitsuyma,S.,
          Antonarakis,S.E., Minoshima,S., Shimizu,N., Norisiek,G.,
          Hornischer,K., Barandt,P., Schafte,W., Schoen,O., Desario,A.,

```

```

TITLE Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
JOURNAL Hennig,S., Rieseemann,L., Dagand,E., Wehrmeier,S., Borzym,K.,
         Gardiner,K., Nizetic,D., Francis,F., Lehnach,H., Reinhardt,R. and
         Vaspou,M.L.
         Direct Submission
         Submitted (10-APR-2000) to the DDBJ/EMBL/GenBank databases. The
         Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic
         Sciences Center, Human Genome Research Group * Institute of
         Molecular Biotechnology, Genome Analysis * Keio University School
         of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome
         Analysis * Max-Planck Institute for Molecular Genetics (addresses
         see below)
         On May 30, 2000 this sequence version replaced gi:7717398.
         The chromosome 21 mapping and sequencing consortium consisting of
         * RIKEN Genomic Sciences Center, Human Genome Research Group, *
         Sagami-hara 228-8555, Japan,
         * e.mail: hattori@gsf.riken.go.jp
         * URL: http://hgp.gsc.riken.go.jp/
         and
         * Institute of Molecular Biotechnology, Genome Analysis, *
         Beutenbergstrasse 11, D-07745 Jena, Germany,
         * e.mail: gscj-submit@genome.imb-jena.de
         * URL: http://genome.imb-jena.de/
         and
         * Keio University School of Medicine, Molecular Biology, * Tokyo
         160-8582, Japan,
         * e.mail: nshimizu@med.keio.ac.jp
         * URL: http://www.dmp.med.keio.ac.jp/
         and
         * GBF, Dept. of Genome Analysis,
         * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
         info.genome@gbf.de
         * URL: http://genome.gbf.de/
         and
         * Max-Planck Institute for Molecular Genetics,
         * Inestrasse 73, D-14195 Berlin, Germany,
         * e.mail: info-chr21@molgen.mpg.de
         * URL: http://chr21.fz-berlin.mpg.de/
         AL163291: Submitted (10-Apr-2000).
         Location/Qualifiers
         source
           1. 340000
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             <1..44256
             /note="Accession No. AP001621"
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="21"
             /map="21q22.3"
             /clone="KB1342D7, 5' partial"
             /clone.lib="Keio BAC library"
             <1..124582
             /note="Accession No. AP001622"
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="21"
             /map="21q22.3"
             /clone="KB1430A10, 5' partial"
             /clone.lib="Keio BAC library"
             join(<310..374,9668..9604)
             /gene="PREDA6"
             /note="exon prediction only"
             310..9804
             /gene="PREDA6"
             310..374
             /gene="PREDA6"
             /number=1
             complement(626..705)
             /note="LIMC4"
             /rpt_family="LINE/L1"
             /rpt_type=DISPERSED
             706..1048

```

[illegible]

RESULT 5
AX017268/c
LOCUS AX017268 834 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 19 from Patent WO947669.
ACCESSION AX017268
VERSION AX017268.1 GI:10042186
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences from tissue of breast tumors
JOURNAL Patent: WO 947669-A 19 23-SEP-1999; DAHL EDGAR (DE); HINZMANN
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
source 1..834
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 173 a 241 c 249 g 171 t
ORIGIN

Query Match 56.7%; Score 516.8; DB 6; Length 834;
Best Local Similarity 92.3%; Pred. No. 6,5e-137;
Matches 596; Conservative 0; Mismatches 2; Indels 48; Gaps 3;

QY 266 aaaaatccttgatgacgtcagctcctcttagagagctcttctgacctgaagccttg 325
Db 834 AAAATATCCTTGATGATGACATGACATGACATGACATGACATGACATGACATG 775

QY 326 gcagactcctcccttgacacccctccgacctcccaagacgacgacgacgacgacgac 385
Db 774 GCAGACTCTCCCTGACACACCTCCCTCCCTGCT---CCACGACGACGACGACGACGAC 718

QY 386 cctcagaagatcctcagcagaagaactgtcctcgtggtgagacatgagaccttatcgt 445
Db 717 CCTCAGAAATCTCAGGACGAGAACTGCTCCTGAGTGTGAGTGTGAGTGTGAGTGTG 658

QY 446 taagacatcagctcagcatatgaacttcagcagaagcgtctgcggagcaaaagac 505
Db 657 TAAGACATCAGCTCCGATATGATGACTTTCAGCAGAACGCTTCCGAGCAAAAGGAC 598

QY 506 agaaaagctgagatgaacagtgctcgtgacgaatcacagccgagcaaggtgtcgcagc 565
Db 597 AGAAAAGCTGAGATGAACAGTGTGCTGACCAATCACAGCCGGGCAAGGCTGCCAGC 538

QY 566 ctgcacatccccgcgagcgagcgtgagtgctcgtcagaagtgtctgtctcctgc 625
Db 537 CTCGCATCCCCCGCGCGGAGCTGAGTGTGCTCAGAAAGTGTGCTTGTCTTCTGC 478

QY 626 agggcttgaacaaccaagcactccagagatcccttgatcaaaagcagcagcccggttg 685
Db 477 AGGGGCTTGAACAACCAAGCAGCTCCAGGATCTTGAGTCAAAAGCAGCAGCCCGGTTG 418

QY 686 ttgacatccttgggggtgacatgggttagcgcgactccacccctgtctgtcgtgacg 745
Db 417 TTGACATCTTGGGGGTGACATGGGGGTAGCCGACGTCCACCTGTCTTGGTGGCAGC 358

QY 746 gcaacactgttgcagctgtccagacaagccctgtcagctgcagagccctgtcgtgg 805
Db 357 GCACACTGTGTTTGA-----G 342

QY 806 acagggccacgtactctcagcagaagctgagagacagcaagccagacagccagc 865
Db 341 ACAGGGCCACGTAC-TCCTCAGCAGAGCTGAGAGACAGCAAGCCAGACAGCCAGC 283

QY 866 atgcagagcgtctgagcagcatgaccacgctggtgtcgcggagcgc 911
Db 282 ATGCAGAGCGCTCTGCGAGCATGACACCGTGGGCTCCGGAGACGC 237

RESULT 6
AX018086/c
LOCUS AX018086 834 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 40 from Patent WO9946374.
ACCESSION AX018086
VERSION AX018086.1 GI:10042537
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences from prostate tumour tissue
JOURNAL Patent: WO 9946374-A 40 16-SEP-1999; DAHL EDGAR (DE); HINZMANN
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
source 1..834
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 173 a 241 c 249 g 171 t
ORIGIN

Query Match 56.7%; Score 516.8; DB 6; Length 834;
Best Local Similarity 92.3%; Pred. No. 6,5e-137;
Matches 596; Conservative 0; Mismatches 2; Indels 48; Gaps 3;

QY 266 aaaaatccttgatgacgtcagctcctcttagagagctcttctgacctgaagccttg 325
Db 834 AAAATATCCTTGATGATGACATGACATGACATGACATGACATGACATGACATG 775

QY 326 gcagactcctcccttgacacccctccgacctcccaagacgacgacgacgacgacgac 385
Db 774 GCAGACTCTCCCTGACACACCTCCCTCCCTGCT---CCACGACGACGACGACGACGAC 718

QY 386 cctcagaagatcctcagcagaagaactgtcctcgtggtgagacatgagaccttatcgt 445
Db 717 CCTCAGAAATCTCAGGACGAGAACTGCTCCTGAGTGTGAGTGTGAGTGTGAGTGTG 658

QY 446 taagacatcagctcagcatatgaacttcagcagaagcgtctgcggagcaaaagac 505
Db 657 TAAGACATCAGCTCCGATATGATGACTTTCAGCAGAACGCTTCCGAGCAAAAGGAC 598

QY 506 agaaaagctgagatgaacagtgctcgtgacgaatcacagccgagcaaggtgtcgcagc 565
Db 597 AGAAAAGCTGAGATGAACAGTGTGCTGACCAATCACAGCCGGGCAAGGCTGCCAGC 538

QY 566 ctgcacatccccgcgagcgagcgtgagtgctcgtcagaagtgtctgtctcctgc 625
Db 537 CTCGCATCCCCCGCGCGGAGCTGAGTGTGCTCAGAAAGTGTGCTTGTCTTCTGC 478

QY 626 agggcttgaacaaccaagcactccagagatcccttgatcaaaagcagcagcccggttg 685
Db 477 AGGGGCTTGAACAACCAAGCAGCTCCAGGATCTTGAGTCAAAAGCAGCAGCCCGGTTG 418

QY 686 ttgacatccttgggggtgacatgggttagcgcgactccacccctgtctgtcgtgacg 745
Db 417 TTGACATCTTGGGGGTGACATGGGGGTAGCCGACGTCCACCTGTCTTGGTGGCAGC 358

QY 746 gcaacactgttgcagctgtccagacaagccctgtcagctgcagagccctgtcgtgg 805
Db 357 GCACACTGTGTTTGA-----G 342

QY 806 acagggccacgtactctcagcagaagctgagagacagcaagccagacagccagc 865

```

|||||
Db 341 ACAGGCCACGTAC-TCTCAGCAGACTGGAGAGACAGCAAGCCAGACCCAGC 283
QY 866 atgagagcgtctgagcagcatgacacccgtggtcccggaagc 911
Db 282 ATGACAGAGCGCTTGAGCAGCATGACACCGTGGCTCCGGAGCGC 237

RESULT 7
AX193486/c AX193486 480 bp DNA PAT 15-AUG-2001
LOCUS Sequence 1053 from Patent WO0149716.
DEFINITION AX193486
ACCESSION AX193486
VERSION AX193486.1 GI:15211437
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,
AUTHORS Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.
TITLE Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use
JOURNAL Patent: WO 0149716-A 1053 12-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source location/Qualifiers
1..480
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 78 a 154 c 141 g 107 t
ORIGIN
Query Match 44.2%; Score 403; DB 6; Length 480;
Best Local Similarity 91.2%; Pred. No. 2.5e-104;
Matches 468; Conservative 0; Mismatches 0; Indels 45; Gaps 2;
QY 399 caggcagaagaactgtctcgtgggtgagagcttgagaccttattcgttaagacatcagc 458
Db 480 CAGGACGAGAACTGCTCTCGGTGGAGCATGGGACCTTTATTCGTAACATCAAGC 421
QY 459 tccagatatgaacttccagcagaagcgctgcccggagcaaaaggagacagaagaactgaga 518
Db 420 TCCAGATATGAACTTTCAGCAGAGCGCTTCCGGAGCAAAAGGACGAAAGCTGAGA 361
QY 519 tgaacagtgcctgagcagaatcacagccggcgaagggtgctcgagctcgatccccc 578
Db 360 TGAACAGTGCCTGGCAGCAATCAACAGCCGGGCAAGGGTGTCCAGCCTCCATCCCCG 301
QY 579 gccgggggagcgtggaagtgctcagaaggtgcatctgtctctgcaagggttgaaac 638
Db 300 GCCGGGGGACACTGGAGGTGCTCAGAAAGGTGCAATTCGCTTCGACAGGGGCTTGAAC 241
QY 639 accaaggaactccaggagatcctgaggtcaaaagcagaccccggtgtgttgcactcttg 698
Db 240 ACCAAGGACATCCAGGGATCTCTGAGTCAAAAGCAGACGCCGGTGTGTGCACTCTGG 181
QY 699 ggggtgacatgggggtgagcgcagtcacacctgtctcttggtgctgcaaggacactgtttg 758
Db 180 GGGGTGACATGGGGGTAGCGGAGTCCACCCCTGCTTGCTGGCAGCGCACATGCTGTTG 121
QY 759 caggtgtccagacaagacctgtcagctgcagaagccctgtcgtggaagaagcccaagta 818
Db 120 CA-----GACAGGCCACAGCTA 105
QY 819 ctctctcagcagagctgagagagacagagccagagccacagcatgagagcgctc 878
Db 104 C-TCTCTAGCAGAGCTGGAGAGACAGCAAGGCCAGAGACCCAGCATGAGAGCGGTC 46
QY 879 tggcagcatgacacacgtgtggtctccgggagcg 911
Db 45 TGGCAGCATGACACCGTGGCTCCGGAGCGC 13

```

```

RESULT 8
HUMPIBX/c HUMPIBX 480 bp mRNA PRI 07-JAN-1995
LOCUS Human secretory protein (Pl.B) mRNA, complete cds.
DEFINITION 115203
ACCESSION 115203.1 GI:402482
VERSION 115203.1 GI:402482
KEYWORDS secretory protein.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE Hauser,F., Poulsom,R., Chinery,R., Rogers,L.A., Hanby,A.M.,
AUTHORS Wright,N.A. and Hoffmann,W.
TITLE hpl.B, a human P-domain peptide homologous with rat intestinal
trefoil factor, is expressed also in the ulcer-associated cell
lineage and the uterus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (15), 6961-6965 (1993)
MEDLINE 93348192
FEATURES
source location/Qualifiers
1..480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/issue_type="Intestine, stomach, uterus"
mRNA 1..460
/gene="Pl.B"
/note="minor transcript"
mRNA 1..480
/gene="Pl.B"
/note="major transcript"
1..480
/gene="Pl.B"
36..98
/gene="Pl.B"
36..278
/gene="Pl.B"
/codon_start=1
/product="secretory protein"
/protein_id="AA59981.1"
/db_xref="GI:402483"
/translation="MAARALCMGLVLTALSSSAEEYVGLSANQCAVPAKDRVDCGY
PHYTPKRNRRGCGFDSIRIPGVWCFPIDEACT"
437..442
/gene="Pl.B"
polya_signal
BASE COUNT 78 a 154 c 141 g 107 t
ORIGIN
Query Match 44.2%; Score 403; DB 9; Length 480;
Best Local Similarity 91.2%; Pred. No. 2.5e-104;
Matches 468; Conservative 0; Mismatches 0; Indels 45; Gaps 2;
QY 399 caggcagaagaactgtctcgtgggtgagagcttgagaccttattcgttaagacatcagc 458
Db 480 CAGGACGAGAACTGCTCTCGGTGGAGCATGGGACCTTTATTCGTAACATCAAGC 421
QY 459 tccagatatgaacttccagcagaagcgctgcccggagcaaaaggagacagaagaactgaga 518
Db 420 TCCAGATATGAACTTTCAGCAGAGCGCTTCCGGAGCAAAAGGACGAAAGCTGAGA 361
QY 519 tgaacagtgcctgagcagaatcacagccggcgaagggtgctcgagctcgatccccc 578
Db 360 TGAACAGTGCCTGGCAGCAATCAACAGCCGGGCAAGGGTGTCCAGCCTCCATCCCCG 301
QY 579 gccgggggagcgtggaagtgctcagaaggtgcatctgtctctgcaagggttgaaac 638
Db 300 GCCGGGGGACACTGGAGGTGCTCAGAAAGGTGCAATTCGCTTCGACAGGGGCTTGAAC 241
QY 639 accaaggaactccaggagatcctgaggtcaaaagcagaccccggtgtgttgcactcttg 698
Db 240 ACCAAGGACATCCAGGGATCTCTGAGTCAAAAGCAGACGCCGGTGTGTGCACTCTGG 181

```


QY	659	gagtgaaacttggggatgagccgagcttcaaccccttcccttgcttgcttgcaacgagcaacttggtg	758
Db	180	GGGTACATATGGGGTGAAGCCGACAGTCCACCTCTCTTGCTGGCTGGACAGGACACTGGTTTG	121
QY	759	caagctgtcccaagacaacacccctgttcagctgtccagagcccttctgtgacagagcccaagta	818
Db	120	CA-----	105
QY	819	cttctcaagcagaagcttgagagacagaagccagagaccagcccgatgtcagaagcgtc	878
Db	104	C-TCTTCAGCAAGAGTGTGGAGGACACCAAGGCCAGGACGCCCCAGCATGACAGAGCGCTC	46
QY	879	tggcagccatgacccaccgctggctcgcggagac	911
Db	45	TGGCAGCCATGACACCGTGGCTCCGGGACGC	13
RESULT	9		
LOCUS	195749	570 bp	DNA
DEFINITION	Sequence 13 from patent US 5733748.		PAT
ACCESSION	195749		
VERSION	195749.1	GI:3940219	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 570)		
AUTHORS	Yu, G. and Rosen, C.		
TITLE	Colon specific genes and proteins		
JOURNAL	Patent: US 5733748-A 13 31-MAR-1998;		
FEATURES	Location/Qualifiers		
source	1..570		
	/organism="unknown"		
BASE COUNT	126 a	171 c	160 g
ORIGIN		113 t	
Query Match	33.5%;	Score 305.4;	DB 6; Length 570;
Best Local Similarity	85.7%;	Pred. No. 2.3e-76;	
Matches 431;	Conservative	0;	Mismatches 16; Indels 56; Gaps 6;
QY	413	tgtcttcggttgagcatgagacctaatttcgttaagaacatcagctccagatgaact	472
Db	552	TTTTTTGGGTGGACATGGGAC-----TACGTTAAACATCAGGCTCCAGATATGAACT	499
QY	473	ttcagcagaagcgtctgcgggagacaaagggacagaaaagctgagatgaacagctgttg	532
Db	498	TTTCAGCAAGACGGCTT-CCGGGAGCAAGAGGACAGAAAAAGCTGAGATGAACAGTCCCTGG	440
QY	533	cagcaatcaagccgggagaaaggtgtctcgaagccctcgatccccccgggc-gggggcagct	591
Db	439	CAGCATACACAGCCCGGCAAGGgtctcGAGcctTcGATCCCGCGCCGGGGGcAGCT	380
QY	592	ggaagtgctcctaagaagtgcatcttgcctt---ccctcagggggtctgtaaacacaaagcac	648
Db	379	GGAGTGGCTTACGAAGTGGCATTCCTGCTTCCCTTCACAGGGCTTGAACACCAAGGCAC	320
QY	649	tccagggatcctgtgagttcaaaagcagcagcccccgttgttgcacctccttgagggtgacatg	708
Db	319	TTCAGGGATCTCTGGAGTCAAGACAGACACCCCGGTTGTTCACCTCCTTGGGGGTGACATG	260
QY	709	ggagtagccgaagctccaccctgtccttcttgcttggcagccgacacactgtttgagcgtgcc	768
Db	259	GGGGTAGCCGACGTCCACCTGTCTTCTTGGCTGGACGGGACACTGGTTTGA-----	206
QY	769	agacaaagccctgttcagctgcagagcccttctgttgcagagcagcccaacttacttctcagc	828
Db	207	-----GACAGGCCACGATAC-TCTTCACG	185
QY	829	agaagctggagacagcaagccagagaccagccccaagcatgcagagcgtcttgcagcat	888

Db	134	AGAGCTGGAGGACACCAAGCCAGGACGCCAGCATGACAGAGCGCTGTGGACGCAT	125
Qy	889	gaccacccgtgggtcccggaagc 911 	
Db	124	GACCACCGTGGGCTCCGGTACGC 102	
RESULT	10		
LOCUS	HUMTREFAC	398 bp	mrna PRI 26-JUL-1993
DEFINITION	Human intestinal trefol Factor mRNA, complete cds.		
ACCESSION	L08044		
VERSION	L08044.1	GI:307520	
KEYWORDS	intestinal trefol factor; trefol protein.		
SOURCE	Homo sapiens adult colon mucosa cDNA to mRNA.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Podolsky,D.K., Lynch-Devaney,K., Stow,D.L., Oates,P., Murgue,B., Debeaumont,M., Sands,B.E. and Mahida,Y.R.		
TITLE	Identification of human intestinal trefol factor: goblet cell-specific expression of a peptide targeted for apical secretion		
JOURNAL	J. Biol. Chem. 268, 6694-6702 (1993)		
MEDLINE	93203271		
FEATURES			
source	Location/Qualifiers		
	1..398	/organism="Homo sapiens"	
	/db_xref="taxon:9606"		
	/tissue_type="colon mucosa"		
	/dev_stage="adult"		
	2..226	/codon_start=1	
	/product="intestinal trefol factor"		
	/protein_id="AAA36766.1"		
	/db_xref="GI:307521"		
	/translation="MGLVYIALSSSSAEYVGLSANQCAVPAKDRDVGCPHYTPKE CNRGCGFSRIRIEVPKFXPLRTKTCIF		
CDS	polya_site 69 a 123 c 111 g 95 t		
BASE COUNT	398		
ORIGIN			
Query Match	29.0%; Score 264.4; DB 9; Length 398;		
Best Local Similarity	91.6%; Pred. No. 1.2e-64;		
Matches 317; Conservative	0; Mismatches 16; Indels 13; Gaps 3;		
Qy	419	cgagtgagagcattgaacttattcgttaagacatcaggtccagatatgaacttcaag 478 	
Db	398	CGGGTGGAGCATGGGACTTTATTCGTTAAGACATCAGAGCTCCACATATGAACCTTCACG 339 	
Qy	479	agaagcgtctgcgggagcaaaagagacagaaagcgtgagatgaaagtgctgcgaagca 538 	
Db	338	AGAA-AGCTTGCCTGGGAGCAAAAGGACGAAAAACTGAGATGAACAGTGCCTGGCAGCAA 280 	
Qy	539	tcaagccgggcaagagtgctcccgagctcgatcccccgccgggggcaacttgagtg 598 	
Db	279	TCACAGCCGGGCAAGGGGTCTAG-----CCTGCATCCAGGGGACCTGAGGAGTG 229 	
Qy	599	cctcgaaggtgcatcttcgtcctc---tgcagggggttgaaacaccgaagcagctccag 655 	
Db	228	CCTCAGAGGATGCAATTCCTGCTTCTGATCAGGGGCTTGAAACACACATCCACAGG 169 	
Qy	656	atccttgagatcaaaagcagagccccggtgtgttgcactccttgagggtgagatggggtag 715 	
Db	168	ATCTCGGAGTCAAAAGACACACCCCGGTTGTTGCACTCTTGGGGGTGACATGGGGGTAG 109 	
Qy	716	ccgagatccacccgtctccttgcttgagcagcgacacagtgattggag 761 	
Db	108	CCGCACTCACCTCTGTCTTGCCGGCCGACAGGCACACTGATTGGCAG 63 	
RESULT	11		

AXI06398/c	AXI06398	250 bp	DNA	PAT	30-APR-2001
LOCUS	Sequence	179 from Patent WO0125272.			
DEFINITION	AXI06398				
ACCESSION	AXI06398.1	GI:13922077			
VERSION					
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 250)				
AUTHORS	Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.				
TITLE	Compositions and methods for therapy and diagnosis of prostate cancer				
JOURNAL	Patient: WO 0125272-A 179 12-APR-2001;				
	CORIXA CORPORATION (US)				
FEATURES	Location/Qualifiers				
source	1..250				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	55 a	75 c	63 g	57 t	
ORIGIN					

Query Match	25.9%	Score 236.2	DB 6	Length 250
Best Local Similarity	98.8%	Pred. No. 1.4e-56		
Matches 238	Conservative 0	Mismatches 3	Indels 0	Gaps 0

QY	413	tgctcccggtgagcaatggaacctattcgttaagaacaatgaagctccagatcatgaact	472
	241	TTTTTTGGGTGAGCATGGGACCTTTATTTCGTTAAACACTCAGGCTCCAGATATGAACT	182
QY	473	ctcagcaagaacgctctgcgaggaacaaagagacagaaaagctgagatgaaacagctgcctg	532
Db	181	TTTCAGCGAAGCGCTTCCCGGGAAGCAAGGAGCAAGAAACCTGAGATGAACTAGTGCCTGG	122
QY	533	caacaaatcacagccgggacaagagtgtgcacgaagctgcaccccccgagcgaggagcaagt	592
Db	121	CAGCAATCACAGCCGGGCAAGGGGTGCTCCAGACCTCCCAATCCCGGCGCGGGGCAAGCTG	62
QY	593	gaagtgctctagaagaatgtcaattctgcttccctgcaagggcttgaaacaaccaagcaatcca	652
Db	61	GAGTGCTCTCAGAAAGTGCAATTGCTGCTCTTCGACGAGGGCTTTGAACACCAAGGCACATCCA	2
QY	653	g 653	
Db	1	G 1	

RESULT 12	AXI40689/c	LOCUS	AXI40688	250 bp	DNA	PAT	31-MAY-2001
DEFINITION	Sequence 179 from Patent WO0134802.	AXI40689					
ACCESSION	AXI40689						
VERSION	AXI40689.1	GI:14280800					
KEYWORDS	.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 250)						
AUTHORS	Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y., Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H., Skelky,Y.A. and Wang,A.						
TITLE	Compositions and methods for the therapy and diagnosis of prostate cancer						
JOURNAL	Patent: WO 0134802-A 179 17-MAY-2001;						
FEATURES	CORIXA CORPORATION (US)						
SOURCE	Location/Qualifiers						
	1..250						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
BASE COUNT	55 a 75 c 63 g 57 t						

```

ORIGIN
Query Match      25.9%      Score 236.2;  DB 6;  Length 250;
Best Local Similarity 98.8%      Pred. No. 1.4e-56;
Matches 238;  Conservative 0;  Mismatches 3;  Indels 0;  Gaps 0;

```

QY	413	tgctctcgggagggagcaltggagacctatcgttgaagaacacagccccaagatgacct	472
Db	241	TTTTTTGGGGGGAGCATGGGACCTTTATTTCGTTAAACATCAGGCTCCAGATATGAACT	188
QY	473	ctcagcagaagcgccttgccgggagcaaaaggacagaaaagctgagatgaaacgtgacctg	532
Db	181	TTTCAGCAGAAACCCCTTCGCCGGAGCAAAAGGACAGAAACCTGAGTAGAACAGTGCCTGG	122
QY	533	cagcaatcacagcccgaggcaagggtgctccgagcctgcacatcccccggccgggggacgtg	592
Db	121	CAGCAATTCACAGCCGGGCGAAGGGTGTCCTCGAGCCTCCATCCCCCGCGGGGACAGCTG	62
QY	593	gaggtgacctcagaaggtgcaatcttcgtcttcctgcagagggtctgaaacacaaaggaaccca	652
Db	61	GAGTGCCTCAGAAAGGTGCATTTCTGCTCTCTCGCAGGGGCTTTGAATACCAAGGACATCTCA	2

OY	653 g	653 z	
	1		
Db	1 G	1	
RESULT	13		
LOCUS	AXI06296/c		
DEFINITION	AXI06296 248 bp	DNA	PAT
ACCESSION	Sequence 77 from Patent WO0125272.		
VERSION	AXI06296		
KEYWORDS	AXI06296.1	GI:13921982	
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:		
AUTHORS	Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.		
TITLE	1 (bases 1 to 248)		
JOURNAL	Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.		
FEATURES	Compositions and methods for therapy and diagnosis of prostate		
source	Patent: WO 0125272-A 77 12-APR-2001;		
	CORIXA CORPORATION (US)		
	Location/Qualifiers		
	1..248		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	55 a	73 c	63 g 57 t
ORIGIN			

	Query Match	24.2%	Score 220.6	DB 6	Length 248
	Best Local Similarity	97.5%	Pred. No. 4.3e-52		
	Matches 235	Conservative 0	Mismatches 4	Indels 2	Gaps 1
QY	413	tgtcctcgggtcggtgcacatggaaccttla	tcgtttaagacatcaaggtccagatatgaact	472	
Db	239	TTTTTTCGGTGGAGACATGGGACCTTTATTCGTATTAAGACATCAGCGCTCCAGATATGAACT		180	
QY	473	ttcgcgcgaagaacgcgtttgcgcgggagcgaagaaggacgaaatcgtgaatgaacagttgcctcgg		532	
Db	179	TTTCAGCGAAGCGCTTGCGCGGGAGCAAAAGGAGCAAAAGCTGAATCAATACAGTCCCTGG		120	
QY	533	cagcaatcacagccgggcggaaggttgtctcacgagccttcgataccccggcgcgggagtcgagctg		592	
Db	119	CAGCAATCACAGCCGGGCGAAGGTGCTCCGAGGCGCTCCGATCCCGG--CGGGGCGAGCTGG		62	
QY	593	gaagtgctcacaagaagtgtgcatlcttgctcttcgtcaggggtctgaaacaccaaagtcactcca		652	
Db	61	GAGGTGCTTCGAAAGGTGCATTTCGCTTCGACAGGGGCTTGAAACACCAAGGCGACTCCA		2	

```
QY 653 g 653
Db 1 G 1

RESULT 14
LOCUS AX140587 248 bp DNA PAT 31-MAY-2001
DEFINITION Sequence 77 from Patent WO0134802.
ACCESSION AX140587
VERSION AX140587.1 GI:14280705
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
Skelly,Y.A. and Wang,A.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0134802-A 77 17-MAY-2001;
CORIXA CORPORATION (US)
FEATURES
source location/Qualifiers
1. 248
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 55 a 73 c 63 g 57 t
ORIGIN

Query Match 24.2%; Score 220.6; DB 6; Length 248;
Best Local Similarity 97.5%; Pred. No. 4.3e-52;
Matches 235; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

QY 413 tgcctcgggtgagcagtgacgttatttcgttaagaacacagctcagatagaact 472
Db 239 TTTTTCGGGTGAGCATGGGACCTTTATTCGTTAAGACATCAGGCTCCAGATATGAACT 180
QY 473 tttagcagaagagccttcgcggagagcaagaagagcaaaactgaatgaacagtcgtcg 532
Db 179 TTGAGCAGAAAGCCCTTGCCGGAGCAAGAGACAGAAAGCTGAGAGAGAGGCTGG 120
QY 533 cagcaatcacagccgggcaagggtgctcgcagcctcgcaccccgccggcgaggagcagctg 592
Db 119 CAGCAATCACAGCCGGGCAAGGAGGTGCTCCGAGCTCGCATCCCGG--CCGGGGCAGCTG 62
QY 593 gaagtgctccagaaggtgcaatctgcttcctcgcagggaggttgaacaacaaggcaactca 652
Db 61 GAGGTGCTCCAGAGGTGCAATCTGCTTCTGCTGAGGGGCTTGAACACCAAGGCACACTCA 2
QY 653 g 653
Db 1 G 1

RESULT 15
LOCUS HSHITF3 320 bp DNA PRI 20-DEC-1995
DEFINITION Human Intestinal Trefol Factor (HITF) gene, exon 3 and complete
cds.
ACCESSION U25657
VERSION U25657.1 GI:940944
KEYWORDS
SEGMENT 3 of 3
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Seib,T., Dooley,S. and Welter,C.
TITLE Characterization of the genomic structure and the promoter region
```

```
of the human intestinal trefol factor
Biochem. Biophys. Res. Commun. 214 (1), 195-199 (1995)
MEDLINE 95398634
REFERENCE 2 (bases 1 to 320)
AUTHORS Seib,T., Dooley,S. and Welter,C.
TITLE Direct Submission
JOURNAL Submitted (24-APR-1995) Thomas Seib, Inst. fuer Humangenetik,
Universitaet des Saarlandes, Oskar-Orth-Strasse, Homburg 66421,
Germany

FEATURES
source location/Qualifiers
1. 320
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
join(U25654.1:<931..1012,U25656.1:295..441,130..320)
/gene="HITF"
join(U25654.1:931..1128,U25656.1:1..730,1..320)
/gene="HITF"
join(U25654.1:931..1012,U25656.1:295..441,130..143)
/gene="HITF"
/note="secretory protein"
/codon_start=1
/product="intestinal trefol factor"
/protein_id="AAA83628.1"
/db_xref="GI:940946"
/translation="MAARALMLGLVLALLSSSAEEYVGLSANOCAPPAKRDVDCGY
PHYTPKECNRRCCFDSPRIPGVPCFKPLQEAECTF"
<1..129
/gene="HITF"
/number=2
130..320
/gene="HITF"
/number=3
BASE COUNT 57 a 97 c 92 g 74 t
ORIGIN

Query Match 21.4%; Score 195; DB 9; Length 320;
Best Local Similarity 100.0%; Pred. No. 9.3e-45;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 gtagcatggacaccttatttcgttaagaacatcaggtccagatagtgaacttcagcgaag 483
Db 320 GGAGCATGGGACCTTTATTCGTTAAGACATCAGGCTCCAGATATGAACTTTCAGCAAG 261
QY 484 cgcttcgggagcaagagagcaaaagctgagatgaaacagtgctggcaagcaatcaca 543
Db 260 CGCTTCCGGGAGCAAGGACAGAAAAGCTGAGATGAAACAGTGCCTGGCAGCAATCACA 201
QY 544 gccgggcaagggtgctcgcagcctcgcaccccgccggcgaggagcagctggaagtgcctca 603
Db 200 GCCGGGCAAGGAGTGTCCGAGCTCCGATCCCGCCGGCGGAGGAGCTGAGAGTGCTCA 141
QY 604 gaagtgcatcttcgc 618
Db 140 GAAGTGCAATCTGC 126
```

Search completed: March 30, 2002, 12:59:16
Job time: 21438 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 07:10:03 ; Search time 519.52 Seconds
(without alignments)
1503.356 Million cell updates/sec

Title:

US-09-867-034-1

Perfect score:

1 ttttttttttcgtgtttt.....cacccgtggtccgcggagacg 911

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_1101:*
1: /SID52/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	521.8	57.3	653	AAH34958	Human colon cancer
c 2	516.8	56.7	834	AAZ33462	Human prostate can
c 3	516.8	56.7	834	AAZ33629	Human breast tumor
c 4	403	44.2	480	AAI29499	C8866 determined c
c 5	378.8	41.6	494	AAZ1864	Human breast and o
c 6	305.4	33.5	570	AAZ45887	Human colon specif
c 7	305.4	33.5	570	AAV16675	Polynucleotide seq
c 8	265.4	29.1	403	AAO28363	Human intestinal t
c 9	265.4	29.1	403	AAI88039	Human intestinal t
c 10	265.4	29.1	403	AAZ57156	Human intestinal t
c 11	259.4	28.5	263	AAZ22480	Human gene signatu

c 12	254.8	28.0	594	AAH34957	Human colon cancer
c 13	236.2	25.9	250	AAV61325	Extended cDNA sequ
c 14	236.2	25.9	250	AAV58649	Prostate tumour sp
c 15	236.2	25.9	250	AAH06412	Human immunogenic
c 16	236.2	25.9	250	AAI10171	Human prostate tum
c 17	236.2	25.9	250	AAH93528	Human prostate-spe
c 18	236.2	25.9	250	AAH84842	Human prostate-spe
c 19	236.2	25.9	250	AAH02593	Prostate tumour an
c 20	220.6	24.2	248	AAV61294	CDNA sequence of p
c 21	220.6	24.2	248	AAV58554	Prostate tumour sp
c 22	220.6	24.2	248	AAV06317	Human immunogenic
c 23	220.6	24.2	248	AAI10076	Human prostate tum
c 24	220.6	24.2	248	AAH34433	Human prostate-spe
c 25	220.6	24.2	248	AAH84747	Human prostate-spe
c 26	220.6	24.2	248	AAH02498	Human prostate-spe
c 27	162.8	20.1	458	AAE92349	Prostate tumour an
c 28	171	18.8	277	AAE68431	Bovine mammary tis
c 29	137	15.0	303	AAE68431	Human lung tumour
c 30	134.4	14.8	206	AAE68353	Human colon cancer
c 31	106.4	11.7	441	AAI16424	Human lung tumour
c 32	102	11.2	431	AAO28362	Human prostate can
c 33	102	11.2	431	AAI8038	Rat intestinal tre
c 34	102	11.2	431	AAH57155	Rat intestinal tre
c 35	60	6.6	8241	AAE29647	Plasmid pICMTFEL
c 36	59.2	6.5	5497	AAE29646	Plasmid pICMTFEL
c 37	58.8	6.5	5142	AAE29645	Plasmid pICMTFEL
c 38	54	5.9	563	AAO67728	Human spasmolytic
c 39	54	5.9	727	AAH3849	Human colon cancer
c 40	52.6	5.8	936	AAE58252	Oligonucleotide D1
c 41	52.6	5.8	936	AAE58254	Oligonucleotide D1
c 42	52.6	5.8	936	AAE58257	Oligonucleotide D1
c 43	52.6	5.8	936	AAE58259	Oligonucleotide D2
c 44	52.6	5.8	936	AAE58262	Oligonucleotide D2
c 45	52.6	5.8	938	AAE58255	Oligonucleotide D1

ALIGNMENTS

RESULT 1	
AAH34958/c	
ID	AAH34958 standard; cDNA: 653 BP.
XX	
AC	AAH34958;
XX	
DT	03-SEP-2001 (first entry)
XX	
DE	Human colon cancer antigen encoding cDNA seq ID NO:2040.
XX	
KW	Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW	colorectal carcinoma; ss.
XX	
OS	Homo sapiens.
XX	
PN	MO200122920-A2.
XX	
PD	05-APR-2001.
XX	
PF	28-SEP-2000; 2000WO-US26524.
XX	
PR	29-SEP-1999; 99US-0157137.
XX	
PR	03-NOV-1999; 99US-0163280.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;
XX	
DR	WPI: 2001-235357/24.
XX	
DR	P-PSDB: AAG75553.
XX	
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX	useful for preventing, diagnosing and/or treating colorectal cancers -
XX	

Db 657 TAAGACATCAGGCTCCAGATATGAACCTTTCAGCAGAGACCGCTTGCCGGAGCAAGAGGAC 598
 QY 506 agaaagcttgatgaaacagtgccctgagcaatcacacgcccgaaggtgtctccagc 565
 Db 597 AGAAAAGCTGAGATGACAGTGCCTGGCAGCAATCACAGCCGGGCAAGGCTGCTCCAGC 538
 QY 566 ctgcacatccccggccgggggagctggaaggctgctcagaaagtgtcattctgtctcgc 625
 Db 537 CTCGCATCCCCGGCGGGGAGCTGAGGTGCCTCAGAAAGGTGATTCGCTTCCTCCGC 478
 QY 626 agggagcttgaacacacaaagcagctccaggaatcctggaagtaaaagcagagcccggttg 685
 Db 477 AGGGGCTTTGAAACACCAAGGCACTCCAGGGATCCTGGAGTCAAAGCAGACGCCGGGTGG 418
 QY 686 ttgacatccttgagggtgacatgagggtgagcagctccaccctgtctcttgctgagcag 745
 Db 417 TTGCATCTCTTGGGGGTGACATGGGGGTAGCCGCAATCCACCTGTCTTGAGTGACAG 358
 QY 746 gcaactgtgttgacgtgtgccagacaaagccctgtcagctgcagagccctgtgtgg 805
 Db 357 GCACACTGGTTTGA-----G 342
 QY 806 acagagcccaagctactctctcagcagagctggaagacagcaagccaggaagccagccagc 865
 Db 341 ACAGGCCCAAGCTAC-TCTTCAGCAGAGCTGGAGGACACCAAGGCCAGGACGCCCCAGC 283
 QY 866 atgcagagcgtctgagcagcatcacacgctgggtctccggaagc 911
 Db 282 ATGCAGAGCGCTGTGGCAGCATGACACCGTGGGCTCCGGGAGCC 237

RESULT 3
 AA233629/c
 ID AA233629 standard; cDNA: 834 BP.

AC AA233629;
 DT 08-DEC-1999 (first entry)
 DE Human breast tumour-associated EST 19.
 KW Expressed sequence tag; EST; human; breast; cancer; gene therapy;
 KW treatment; tumour; cytosolic; medicament; ss.
 OS Homo sapiens.
 XX DEL9813839-A1.
 PN 23-SEP-1999.
 PD 20-MAR-1998; 98DE-1013839.
 PE 20-MAR-1998; 98DE-1013839.
 PR 20-MAR-1998; 98DE-1013839.
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
 PA Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
 PI WPI; 1999-528981/45.
 DR P-PSDB; AAY48551.
 DR Human nucleic acid sequences and protein products from tumor breast
 PT tissue, useful for breast cancer therapy -
 XX
 PS Claim 3; 99; 188pp; German.
 CC This invention describes novel human nucleic acid sequences from tumor
 CC breast tissue which have cytosolic activity. The nucleic acid sequences
 CC can be used to produce and isolate full-length gene sequences. They can
 CC be used to express proteins, which can be used as tools to find an
 CC activity against breast cancer. The sequences can be used in sense or
 CC antisense form. They are especially useful for medicaments for gene
 CC therapy to treat breast cancer. AA233611-248617 represents expressed

CC sequence tags described in the method of the invention.
 XX
 SQ Sequence 834 BP; 173 A; 241 C; 249 G; 171 T; 0 other;

Query Match 56.7%; Score 516.8; DB 20; Length 834;
 Best Local Similarity 92.3%; Pred. No. 6,7e-147;
 Matches 596; Conservative 0; Mismatches 2; Indels 48; Gaps 3;

QY 266 aaatatccttgatgacatgacatcctccttaggaagtttctcgtcccttgagccttg 325
 Db 834 AAAATATCCTTGATGACACTCAGCTCCTTAGGGAGTCTTTCTGCGCTTGAGGCTGG 775
 QY 326 gcaactcctccggaacccctccgacctctcccaagcagaagaataaagacaa 385
 Db 774 GCAGACTTCCCTGACACCTCCCGCTGT---CCACAGCAGACAGCAATAATACACAA 718
 QY 386 cctcagaagctcagcagcagaagaaactgtcctcgggtgagcaltggaacttaltcgt 445
 Db 717 CCTGAAAGTCTCAGGACGAGAACTGTCTGGGTGAGCATGGGACCTTATTGCT 658
 QY 446 taagacatcaggtctccagatatgaatttcagcagaagcgttcgagagcaaaaggac 505
 Db 657 TAAGCATCAGGCTCCAGATATGAACCTTCAGCAGAAAGCGCTTGCCGGAGCAAGGAC 598
 QY 506 agaaagcttgatgaaacagtgccctgagcaatcacacgcccgaaggtgtctccagc 565
 Db 597 AGAAAAGCTGAGATGACAGTGCCTGGCAGCAATCACAGCCGGGCAAGGCTGCTCCAGC 538
 QY 566 ctgcacatccccggccgggggagctggaaggctgctcagaaagtgtcattctgtctcgc 625
 Db 537 CTCGCATCCCCGGCGGGGAGCTGAGGTGCCTCAGAAAGGTGATTCGCTTCCTCCGC 478
 QY 626 agggagcttgaacacacaaagcagctccaggaatcctggaagtaaaagcagagcccggttg 685
 Db 477 AGGGCTTTGAAACACCAAGGCACTCCAGGATCCTGGAGTCAAACACACACCCCGGTGG 418
 QY 686 ttgacatccttgagggtgacatgagggtgagcagcagctccaccctgtctcttgagcagc 745
 Db 417 TTGCATCTCTTGGGGGTGACATGGGGGTAGCCGCAATCCACCTGTCTTGAGTGACAG 358
 QY 746 gcaactgtgttgacgtgtgccagacaaagccctgtcagctgcagagccctgtgtgg 805
 Db 357 GCACACTGGTTTGA-----G 342
 QY 806 acagagcccaagctactctcctcagcagagctggaagacagcaagccaggaagccagccagc 865
 Db 341 ACAGGCCCAAGCTAC-TCTTCAGCAGAGCTGGAGGACACCAAGGCCAGGACGCCCCAGC 283
 QY 866 atgcagagcgtctgagcagcatcacacgctgggtctccggaagc 911
 Db 282 ATGCAGAGCGCTGTGGCAGCATGACACCGTGGGCTCCGGGAGCC 237

RESULT 4
 AA129499/c
 ID AA129499 standard; cDNA: 480 BP.

AC AA129499;
 DT 12-OCT-2001 (first entry)
 DE C886P determined cDNA sequence.
 KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 KW immunogenic; gene therapy; vaccine; colonic cancer; ss.
 OS Homo sapiens.
 XX WO200149716-A2.
 XX 12-JUL-2001.
 XX

PF 29-DEC-2000; 2000MO-US355596.
 XX 30-DEC-1999; 9905-0476296.
 PR 10-JAN-2000; 2000US-0480321.
 PR 15-FEB-2000; 2000US-0504629.
 PR 06-MAR-2000; 2000US-0519444.
 PR 19-MAY-2000; 2000US-0575251.
 PR 29-JUN-2000; 2000US-0609448.
 PR 28-AUG-2000; 2000US-0649811.
 XX
 PA (CORI-) CORIXA CORP.
 PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
 PI King GE, Wang T, Jiang Y;
 DR WPI; 2001-441847/47.
 XX
 PT Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer -
 XX
 PS Claim 2; Page 423; 472pp; English.
 XX
 CC The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patients own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and
 CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512
 CC and AAAM24494 to AAAM24523 represent nucleotide and amino acid sequences
 CC given in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 480 BP; 78 A; 154 C; 141 G; 107 T; 0 other;

Query Match 44.28; Score 403; DB 22; Length 480;
 Best Local Similarity 91.2%; Pred. No. 1,9e-112;
 Matches 468; Conservative 0; Mismatches 0; Indels 45; Gaps 2;

OY 399 caagcagcaagactctcccggtgagagatggagacttattcgttaagacatggc 458
 DB 480 CAGGCACGAGAACTGCTCCGCGGAGCATGGACCTTATTCGTAAGACATCAGGC 421
 OY 459 tccagatagacttcagcagaagcgttcgcggagcaagaaggaagaagctaga 518
 DB 420 TCCAGATATGAACTTTCAGCAGAAAGCGCTTGGCGGAGCAAGGAGCAAGAAAGCTAGA 361
 OY 519 tgaacagctgctggcagcaatcaccagcgggcaagggctgcagcctgcaccccg 578
 DB 360 TGAACAGTGTGCTGCGACCAATTCACAGCCGGGCAAGGGTGTCTCCGAGCTGCATCCCCCG 301
 OY 579 ggcgggggagcgtggaggtgcctcaagaagtgatctgctctgtcgaagggttgaaac 638
 DB 300 GCGGGGGGCGCTGGAGGTCTCTCAGAGGTGCAATTCCTCTCCGAGGGGCTTGAAAC 241
 OY 639 accaagagcactccaggaatcctggagctcaagaagcagcagcccggttcactccttg 698
 DB 240 ACCAAGGACATCCAGGAGTCTCTGAGTCAAGCAGCAGCCCGGTTTCTTCCACTCCTTGG 161

OY 699 gggtagacatgggggtagccgcaatccaccctgtccttgctgctggcagcgacactglttg 758
 DB 180 GGGTAGACATGGGGGTAGCCGACAGTCCACCTGTCTTGGCTGGCAGCGCACATGGTTTG 121
 OY 759 cagctgtcccaagacaagaccctgtcagctgcagagagccctgtctggagagggcccgcta 818
 DB 120 CA-----GACAGGCCACAGTA 105
 OY 819 ctctcctcagcagcgttggagagcagaagcagcagaccagccccaatcagcagcctc 878
 DB 104 C-TCTCAGCAGAGCTGGAGAGCAGCAGGCAAGACACCTCCACAGATGCAAGCGCTC 46
 OY 879 tggcagcagcagcagccagcgttggtcgcggagcgc 911
 DB 45 TGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 13

RESULT 5
 AAF21864/c
 ID AAF21864 standard; DNA; 494 BP.
 XX
 AC AAF21864;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Human breast and ovarian cancer associated antigen gene SEQ ID 251.
 XX
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antiviral; antiparasitic; cardiant; immune disorder;
 KW antibacterial; antifungal; autoimmune haemolytic anaemia;
 KW Addison's disease; allergy; diabetes mellitus; Crohn's disease;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200055173-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000MO-US05881.
 XX
 PR 12-MAR-1999; 9905-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-611515/58.
 DR P-PSDB; AAB58961.
 PT
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 XX
 PS Claim 1; Page 673; 1299pp; English.
 XX
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antiviral; antiparasitic; cardiant; immune;
 CC antibacterial; antifungal; autoimmune haemolytic anaemia;
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,

CC proteins, agonists and agonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, Rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.

XX
SO Sequence 494 BP; 121 A; 143 C; 132 G; 97 T; 1 other;

Query Match 41.6%; Score 378.8; DB 21; Length 494;
Best Local Similarity 90.1%; Pred. No. 4.5e-105;
Matches 445; Conservative 1; Mismatches 3; Indels 45; Gaps 2;

Qy 413 tgcctcggggtgagcagtgagccttattcgttaagcattcagcctccagatatgaact 472
Db 450 TTTTTCGGGTGAGACATGGACCTTTATTCGTTAAGACATCAGGCTCCAGATATGAAC 391
Qy 473 ttcaagcaagcgccttcgaggaagcaaggaagcaagcaagctgatatgaagctctg 532
Db 390 TTTCAGCAAAACGCTGCGCGGAGCAAAAGGAGCAAAAGCTGAGATGACAGTCTCTGG 331
Qy 533 cagcaatcacagccgaggaaggtgtctcgagcctcgatcccccgccgagggagctg 592
Db 330 CAGCAATCACAGCCGGGCAAGGGTCTCCGAGCCTCGCATCCCCCGCGGGGAGAGCTG 271
Qy 593 gaggtgctcagaaggtgcatcttgccttcctcgagggagcttgaacaccagaagctcca 652
Db 270 GAGGCGCTTCAGAAAGTGCATTCTCTTCGTCAGAGGGCTTGAAACACCAAGCCACTCCA 211
Qy 653 ggaatcctcgtgagcaaaagcaagcccggtgtgtcactcctctgggggtgacatg99g 712
Db 210 GGGATCCTCGAGTCAAAACAGACAGCCCGGTTTGCACCTCTTGGGGTGACATGGGG 151
Qy 713 tagccagagtcacacctgtccttgctgagcaagcagacatggttgcagctgtccagac 772
Db 150 TAGCCGAGTCCACCCCTGCTTGCTGGTCAGACGACACTGGTTGCA----- 101
Qy 773 aaagccctgtagctgcagagcccttgctggaagggcccaagctacttctctcagag 832
Db 102 -----GACAGGCCACACCTAC-TCTTCACGACAG 76
Qy 833 ctggaagcagcaagcagcagcagcccccagcagcagcagcagcagcagcagcagc 892
Db 75 CTGGAGGAGCAGACAGGCCAGGACGACCCCGACATGACAGATGCTCTGGCAGCCATGACC 16
Qy 893 accgttgagctcgg 906
Db 15 ACCGTGGGCTCCGG 2

RESULT 6
AAT45887/c
ID AAT45887 standard; cDNA: 570 BP.

XX
AC AAT45887;

XX
DT 13-MAR-1997 (first entry)

XX
DE Human colon specific gene CSG8 cDNA full-length clone.

XX
KM Colon specific gene; CSG8; colon cancer; metastasis; diagnosis;

XX
XX gene therapy; ss.

OS Homo sapiens.

XX
FH Key

XX
FT CDS

XX
PN 125..370
XX /*tag= a
XX W09639419-A1.

PD 12-DEC-1996.

XX 06-JUN-1995; 95WO-US07289.

XX 06-JUN-1995; 95WO-US07289.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Yu G;

XX WPI: 1997-043054/04.

XX P-PSDB; AAW06550.

PT Human colon specific genes and their expression products - detection
PT of which, in non-colon tissue samples, can be used as indication of
PT colon cancer metastasis

XX
PS Claim 15; Fig 8; 60pp; English.

CC 13 cDNA clones (AAT45880-92), only 2 of which are full length,
CC correspond to human colon specific genes, designated CSG1, CSG2,
CC etc., that are primarily expressed in tissues derived from the
CC colon. CSG7 and CSG10 show reduced expression in colon cancer
CC cells as compared to that in normal cells; the remaining genes are
CC overexpressed in colon cancer. The full-length sequences can be
CC used to isolate genomic clones including the complete gene. CSG
CC nucleic acids can be used to produce CSG polypeptides (see also
CC AAW06545-53) in transformed host cells, as probes to detect disorders
CC of the colon, partic. colon cancer and colon cancer metastasis, and
CC in gene therapy.

XX
SO Sequence 570 BP; 126 A; 171 C; 160 G; 113 T; 0 other;

Query Match 33.5%; Score 305.4; DB 18; Length 570;
Best Local Similarity 85.7%; Pred. No. 1e-82;
Matches 431; Conservative 0; Mismatches 16; Indels 56; Gaps 6;

Qy 413 tgcctcggggtgagcagtgagccttattcgttaagcattcagcctccagatatgaact 472
Db 552 TTTTTCGGGTGAGACATGGAC-----TACGTTAAACATCAGGCTCCAGATATGAAC 499
Qy 473 ttcaagcaagcgccttcgaggaagcaaggaagcaagcaagctgatatgaagctctg 532
Db 488 TTTCAGCAAAACGCTT-CCGGAGCAAAAGGAGCAAAAGCTGAGATGACAGTCTCTGG 440
Qy 533 cagcaatcacagccgaggaaggtgtctcgagcctcgatcccccgcc-ggggcagct 591
Db 439 CAGCAATCACAGCCGGGCAAGGGTCTCCGAGCCTCGCATCCCCCGCGGGGAGAGCT 380
Qy 592 ggaagtgctcagaaggtgcatctgcct---ccgcaggggtgtgaaccccaagcagc 648
Db 379 GGAGGTGCTTCAGAAAGTGCATTCTGCTTCCCTGTCAGGGGCTGAAACACCAAGGAC 320
Qy 649 tccaaggaatcctgaggtcagaagcagcagcccggtgtgtcagcctctcgggggtacatg 708
Db 319 TTCAGGATCTCTGAGTCAAAACAGACAGCCCGGTTTGTTCATCTCTTGCGGGGTACATG 260
Qy 709 ggggtagcagcagtcacacctgtccttgctgagcagcagcagcagcagcagcagcagc 768
Db 259 GGGGTAGCCGACAGTCCACCTCTGCTTGCTGGTCAGCGACACTGTTGCA----- 206
Qy 769 agacaagcctgtcagctgcagagcccttgctggaagcagcagcagcagcagcagcagc 828
Db 207 -----GACAGGCCACCTAC-TCTTCAGC 185
Qy 829 agagctgagagcagcagcagcagcagcagcccccagcagcagcagcagcagcagcagcagc 888
Db 184 AGAGCTGAGAGCAGCAAGGCCAGGACGACCCCGACATGACAGAGCTCTGGCAGGCAT 125
Qy 889 gaccacgctgggctcgggagcgc 911
Db 124 GACCACGCTGGCTCCGCTACGC 102

Best Local Similarity 91.6%; Pred. No. 1,2e-70;
Matches 318; Conservative 0; Mismatches 16; Indels 13; Gaps 3;

QY 418 tcgggtgagatgagatggagacccttattcgttaagacatcagctccagatataaactttag 477
Db 399 TCGGGTGGAGCATGGGACCTTTATTGTTAAGACATCAGGCTCCAGATATGAATTTTCAG 340
QY 478 cagaagcgttgcggcgagcaaaagagacagaagaactagatgaacatgtcctgacaga 537
Db 339 CAGAA-ACCTTGCCGGAGCAAAAGGACAGAAAACTGATGAACTGCTGCGACACA 281
QY 538 atcaagccggcgagcaagggtgtcgcagcctgcacatcccccgcgcgaggcgatgaggt 597
Db 280 ATCACAGCCGGGCAAGGAGGTGCTCAG-----CCTGCATCCCGAGGGGAGCTGGAGGT 230
QY 598 gcccaagaagtgatctgtcttc---tgcagggtctgaaacccaagcactccag 654
Db 229 GCCTCAGAAAGGTGCAATGTTGCTTCTAGTCAAGGGGCTTGAACACCAAGGCACCTCCAG 170
QY 655 gatcctgagtcacaaagcagcagcccggtgtgtgaccccttggggtgacatggggtga 714
Db 169 GATCCTGGAGTCAAAAGCAGCAGCCCCGGTGTGTGACATCTCTGGGGGTGACATGGGGTA 110
QY 715 gccgagtcacacccctgtccttgcctgagcagcgacactgattgag 761
Db 109 GCCGCAGTCCACCTGTGCTTGGCCGCGACGACACACTGTTTCAG 63

RESULT 9
AAT8039/c
ID AAT8039 standard; cDNA; 403 BP.

AC AAT8039;

DT 27-APR-1998 (first entry)

DE Human intestinal trefoil factor (hTFF) encoding cDNA.

KW Trefoil factor; intestinal; TFF; human; inhiblt; lesion; ulcer;

KW inflammation; cancer; treatment; ss.

OS Homo sapiens.

FX Key Location/Qualifiers

FT CDS 2..226

FT /tag= a

FT /product= "intestinal trefoil factor"

PN MO9738712-A1.

PD 23-OCT-1997.

PF 11-APR-1997; 97WO-US06004.

PR 12-APR-1996; 96US-0631469.

PA (GEHO) GEN HOSPITAL CORP.

PI Podolsky DK;

DR WPI; 1997-526205/48.

DR P-PSDB; AAW27631.

PT Trefoil polypeptide to treat or inhibit lesion formation - useful to
PT treat or protect tissues against wounds, e.g. ulcers, inflammation,
PT cancer and other insults

PS Disclosure: Fig 6; 65pp; English.

XX This cDNA encodes a human intestinal trefoil factor (hTFF). A clone
CC comprising the DNA sequence of rat hTFF is used to identify a cDNA clone
CC encoding this hTFF in a human intestinal cDNA library. The trefoil
CC polypeptide, or a fragment of it can be used to treat or inhibit lesion

CC formation. The TFF polypeptide can be used to treat or protect against
CC eye, skin or gastrointestinal or genitourinary tract wounds, e.g.
CC lesions, ulcers, burns, inflammations or abrasions. It can also be used
CC to treat non-ulcer dyspepsia, gastritis, peptic or duodenal ulcer,
CC gastric cancer, MALT lymphoma, Menetrier's syndrome, gastro-oesophageal
CC reflux disease, Crohn's disease, inflammatory bowel disease or neoplastic
CC cancer. It can be used to protect tissues against injury caused by
CC radiation, bacterial infection, alcohol or drugs. It can also be used
CC for antibody production and diagnosis.

XX Sequence 403 BP; 74 A; 123 C; 111 G; 95 T; 0 other;

Query Match 29.1%; Score 265.4; DB 18; Length 403;
Best Local Similarity 91.6%; Pred. No. 1,2e-70;
Matches 318; Conservative 0; Mismatches 16; Indels 13; Gaps 3;

QY 418 tcgggtgagatgagatggagacccttattcgttaagacatcagctccagatataaactttag 477
Db 399 TCGGGTGGAGCATGGGACCTTTATTGTTAAGACATCAGGCTCCAGATATGAATTTTCAG 340
QY 478 cagaagcgttgcggcgagcaaaagagacagaagaactagatgaacatgtcctgacaga 537
Db 339 CAGAA-ACCTTGCCGGAGCAAAAGGACAGAAAACTGATGAACTGCTGCGACACA 281
QY 538 atcaagccggcgagcaagggtgtcgcagcctgcacatcccccgcgcgaggcgatgaggt 597
Db 280 ATCACAGCCGGGCAAGGAGGTGCTCAG-----CCTGCATCCCGAGGGGAGCTGGAGGT 230
QY 598 gcccaagaagtgatctgtcttc---tgcagggtctgaaacccaagcactccag 654
Db 229 GCCTCAGAAAGGTGCAATGTTGCTTCTAGTCAAGGGGCTTGAACACCAAGGCACCTCCAG 170
QY 655 gatcctgagtcacaaagcagcagcccggtgtgtgaccccttggggtgacatggggtga 714
Db 169 GATCCTGGAGTCAAAAGCAGCAGCCCCGGTGTGTGACATCTCTGGGGGTGACATGGGGTA 110
QY 715 gccgagtcacacccctgtccttgcctgagcagcgacactgattgag 761
Db 109 GCCGCAGTCCACCTGTGCTTGGCCGCGACGACACACTGTTTCAG 63

RESULT 10
AA57156/c
ID AA57156 standard; cDNA; 403 BP.

AC AA57156;

DT 16-OCT-2000 (first entry)

DE Human intestinal trefoil factor cDNA.

KW Human, intestinal trefoil factor; hTFF; antiulcer; antiinflammatory;

KW antimicrobial; cytostatic; gastrointestinal motility enhancer;

KW peptic ulcer disease; inflammatory bowel disease; anticancer;

KW gastrointestinal tract protection; bacterial infection; radiation injury;

OS Homo sapiens.

FX Key Location/Qualifiers

FT CDS 2..226

FT /tag= a

FT /product= "hTFF"

PN US6063755-A.

PD 16-MAY-2000.

PF 07-JUN-1995; 95US-0476705.

PR 02-FEB-1994; 94US-0191352.

PR 13-FEB-1992; 92US-0837192.

PR 25-MAR-1993: 9305-0037741.
 PR 14-FEB-1991: 9105-0655965.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Podolsky DK;
 DR WPI: 2000-364484/31.
 DR P-PSDB; AAY99888.
 XX
 PT New rat intestinal trefoil factor, useful for treating peptic ulcer
 PT diseases, or inflammatory bowel diseases and for protecting the
 PT intestinal tract from injury caused by bacterial infection, or
 PT radiation injury -
 XX
 PS Disclosure; Fig 6; 17pp; English.
 XX
 CC The present sequence encodes human intestinal trefoil factor (hITF).
 CC It was isolated from a human colon cDNA library. The library was
 CC screened using a probe made from a fragment of human cDNA which had been
 CC amplified from human colon library cDNA using probes based on the
 CC sequence of rat intestinal trefoil factor. ITF may be used for the
 CC treatment of peptic ulcers and inflammatory bowel disease, and for
 CC protection of the intestinal tract from injury caused by bacterial
 CC infection or radiation injury. ITF may also be used to produce
 CC monoclonal antibodies for the detection of ITF in an intestinal tissue
 CC or blood serum by indirect immunosay. ITF can also be used to treat
 CC neoplastic cancer.
 CC
 SQ Sequence 403 BP; 74 A; 123 C; 111 G; 95 T; 0 other;
 XX
 XX
 Query Match 29.1%; Score 265.4; DB 21; Length 403;
 Best Local Similarity 91.6%; Pred. No. 1.2e-70;
 Matches 318; Conservative 0; Mismatches 16; Indels 13; Gaps 3;
 QY 418 tcgggtgagcatatggaacattatcgttaagacatagctccagatataaacttcag 477
 DB 399 TCGGGTGGAGCATGAGACCTTATTCTTAAGACATAGGCTCCAGATATGAACTTCAG 340
 QY 478 cagaagagcttgcggagaaagagagagaaagctgagatgaacagctgctgagca 537
 DB 339 CAGAA-AGCTTGGCCGGAGCAAAAGGACAGAAACTGAGATGAACAGTGCCTGGACGA 281
 QY 538 atcacagcggagagaggtgctccagagctcgatcccccggggagcagctgagagt 597
 DB 280 ATCAACAGCCGGGCAAGGAGGTGCTCAG-----CTGCAATCCAGGGGAGCTGGAGGT 230
 QY 598 gctcagaaggtgcatctgctcc---tgcaagggtctgaaacacaaaggaactcag 654
 DB 229 GCGTCAGAAAGGTGATCTGCTCTCTAGTCAGGGGCTTGAACACCAAGGCACTCCAGG 170
 QY 655 gatctgagatcaaaagagcagcccggttgctgactcccttggagggtgacatggagta 714
 DB 169 GATCTTGAGAGCAAAAGAGCAGCCCGGTGTGTTCACTCCTGGGGGTGACATGGGGTA 110
 QY 715 gccgcaatcacccctgctctgctgagcagcagcagctggttcag 761
 DB 109 GCGGCACTCCACCTGTCTTGGCGGCGACGCGACACTGGTTTCAG 63
 RESULT 11
 AAT22480/c
 ID AAT22480 standard; cDNA to mRNA; 263 BP.
 XX
 AC AAT22480;
 XX
 DT 22-AUG-1996 (first entry)
 XX
 DE Human gene signature HDMGS04091.
 XX
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW

KM cell typing; abnormal cell function; ss.
 XX Homo sapiens.
 OS
 XX
 PN W09514772-A1.
 XX
 PD 01-JUN-1995.
 XX
 PF 11-NOV-1994: 94W0-JP01916.
 XX
 PR 12-NOV-1993: 93JP-0355504.
 PA (MATS/) MATSUBARA K.
 PA (OKUBA/) OKUBO K.
 XX
 PI Matsubara K, Okubo K;
 DR WPI: 1995-206931/27.
 XX
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 PS Claim 1; Page 1138; 2245pp; Japanese.
 XX
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T2637 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 CC
 SQ Sequence 263 BP; 47 A; 80 C; 70 G; 64 T; 2 other;
 XX
 XX
 Query Match 28.5%; Score 259.4; DB 16; Length 263;
 Best Local Similarity 98.9%; Pred. No. 6.3e-69;
 Matches 260; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 396 tctcagcagaaagaaactgtcctcgggtgagacatggaaccttattcgttaagacatca 455
 DB 263 TTTCAGGCAAGAAAGAACTGTCTCGGTGAGCATGGACCTTATTCTTAAGACATCA 204
 QY 456 ggtctcagatatgaacttcaagcagaagcgttcgagagcaaaaggacagaagaactg 515
 DB 203 GCGTCCAGATATGAACCTTACAGCAAGAGCGCTTCCGGGAGCAAAAGGCAAAAGCTG 144
 QY 516 agattgaacagtgtcctgcaacaatacacagcggggaaggtgtctcgaagcctgcatccc 575
 DB 143 AGATGAAACAGTGTCTGACACAAATCACAGCCGGGCAAGGAGTGTCTCCAGCTGCATCCC 84
 QY 576 ccgagcggggcagctggaaggtgctcagaagggtgcatctgtctcctcgaagggtctga 635
 DB 83 CCGGCGGGGGGAGCTGGAGGTGCTCTANAAAGMTGATTTGCTTCTCGAGGGGCTTGA 24
 QY 636 aacaacaaaggaactcagaagatc 658
 DB 23 AACACCAAGGCACTCCAGGATC 1
 RESULT 12
 AAH34957/c
 ID AAH34957 standard; cDNA; 594 BP.
 XX
 AC
 XX
 DT
 XX
 DE
 XX
 KW
 KW

XX	AAH34957;
AC	
XX	03-SEP-2001 (first entry)
DT	
XX	
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:2039.
XX	
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW	colorectal carcinoma; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200122920-A2.
PD	
XX	05-APR-2001.
XX	
PF	28-SEP-2000; 2000WO-US26524.
XX	
PR	29-SEP-1999; 99US-0157137.
PR	03-NOV-1999; 99US-0163280.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;
XX	
DR	WPI; 2001-235357/24.
DR	P-PSDB; AAG75552.
XX	
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT	useful for preventing, diagnosing and/or treating colorectal cancers -
XX	
PS	Claim 1; Page 3532; 9803pp; English.

AAH332943, AAH371195 and AAG75314 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent sequences used in the exemplification of the present invention.

N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 594 BP; 122 A; 173 C; 177 G; 116 T; 6 other;

Query Match	28.0%;	Score 254.8;	DB 22;	Length 594;
Best Local Similarity	87.2%;	Pred. No. 2.5e-67;		
Matches 354;	Conservative	1;	Mismatches 3;	Indels 48;
				Gaps 5

QY	506	aggaagctgagtagaacacagtgccctgacaggaatcaacagccgagcaagggctgctcgagc	555
Db	531	ACAAAAAGTGTGAGTGAACAGTGGCTT-GCAGCAATTCACA-CCGGGCAAGGGGTGCTCCGAGC	474
QY	566	ctcgatcccccgcgcgagcgagctggaagtgctgcataagaagtgcattctgctctcgc	625
		:::	
Db	473	YTCGCAATCCCGCCGGCC-GGGGACACTGGAAGTGGCTTAGAAGTGCACTTGTGCTTCTGTC	415
QY	626	aggggcttgtaaacacccaagcagctcccaaggagatccctgagtcacaagcagcagcccggttg	685
Db	414	AGGGGCTTGAACAACCAAGCAGCTCCAGGATCTCTGGAGTCAAAAGCAGCAGCCCGGTTG	355
QY	686	ttgcactctcttggggtgacatgggggtlaagccgacagtcacacccctgtccttgctgctgacag	745

Db	354	TTGGACTCTCTTGGGGGTGATGGGGGAGCCGAGTCCACCCCTGTCTTGGCTGGGACG	2955
Qy	746	ggacacttggttgacgucgtgccagacaagccctgcagctgcagagacccttgcggg	8055
Db	294	GCACACTGGTTTGGC-----G 279	
Qy	806	acagagccacgactctcttcagcagagcttgagagacagacaagcgcagagaccagccagc	8655
Db	278	ACAGGCCCAACGTAC-TCTTCACGACAGCTGGAGGACACAGCAGGACACAGCCACG	220
Qy	866	atcagagagcctctgcagagcatgacacacgctggagcttcgcgagacg 911	
Db	219	ATGCGAAGGGCTCTGGCAGCCATGACCAACCGTGGAGCTCCGGGACG 174	

RESULT 13
AAV61325/C
ID AAV61325 standard; cDNA; 250 BP.

DT , 06-JAN-1999 (first entry)

DE Extended cDNA sequence of prostate tumour clone 1G-4736.

KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.

OS Homo sapiens

PN W09837093-A2

PD 27-AUG-1998.

25-FEB-1998; 98WO-US03492.

PR 09-FEB-1998; 98US-0020956.

PR 01-AUG-1997; 97US-0904804

PA (CORI-) CORIXA CORP

PI Dillon DC, Xu J

WPI; 1998-609886/51.

Polypeptides comprising immunogenic portions of prosta

used in a vaccine for the treatment of prostate cancer

XX FS CLAIM 12, Page 109-110; 130pp; ENGLISH.

cc of a prostate tumour protein. The encoded immunogen, or the DNA itself, the present sequence is a DNA which encodes an immunogenic portion

CC can be used as a vaccine for the treatment of prostate cancer. The DNA CC was identified by analysis of a subtracted cDNA library obtained by

CC subtracting a prostate
CC tissue cDNA library

Sequence 250 BP; 55 A; 75 C; 63 G; 57 T; 0 other;

Query Match	25.9%	Score 236.2	DB 19	Length 250
Best Local Similarity	98.8%	Pred. No. 6.9e-62		
Matches 238; Conservative	0	Mismatches 3	Indels 0	Gaps 0

QY	413	193	ctccgcgggaggaacagggaccttttcgtttaagacatcaagagccacgatagact	472
Db	181	TTTCACCAACAAGCGCTTCCGGGAGACCAAAAGGACGAAAGAGTGTGATGAACACAGTCCGTG	1222	
Db	241	TTTTTTCCGGGTGACCAATGGGACCTTTATTTCGTTAAGACATCAGCGCTCCAGATATGACT	1822	
QY	473	ttcagcagaagcgccttcgcgcggaggaacaaaggacagaaagctgagatgaacagtgccctgg	5322	

QY 533 cagcaatcacagccggcgaaggtgtctccgagcctcgatcccccggccggggagctg 592
 |||||||
 Db 121 CAGCAATACAGCCGGGCAAGGTCCTCCGAGCCTCGCATCCCGCGCGGGGCGAGCTG 62
 QY 593 gaggtgctcagaaggtgtcattctgtctccgcaggggcttgaacaacccaagcactca 652
 |||||||
 Db 61 GAGGTGCTCAGAAAGTGCAATTCTGCTTCGACAGGGGCTTGAAACACCAAGGCACCTCA 2
 QY 653 g 653
 |
 Db 1 G 1

RESULT 14
 AAV58649/C
 ID AAV58649 standard; cDNA: 250 BP.
 XX

AC AAV58649;

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone Ig-4736.

KM Prostate tumour specific gene; human; prostate cancer; detection;

KW therapy; ss.

OS Homo sapiens.

PN WO9837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; 98WO-US03690.

PR 09-FEB-1998; 98US-0904809.

PR 25-FEB-1997; 97US-0806596.

PR 01-AUG-1997; 97US-0904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

PI WPI; 1998-480805/41.

PT Novel human prostate specific tumour protein and fragments - useful

PT for detecting and treating prostate cancers

PS Claim 1; Page 118; 141pp; English.

XX This sequence represents a human prostate tumour specific gene, and can
 CC be used in the method of the invention. The method is for detecting
 CC prostate cancer comprises contacting a biological sample with an agent
 CC able to bind an immunogenic portion of a prostate protein (such as
 CC encoded by this sequence). An antibody which binds to an immunogenic
 CC portion of the prostate protein, and the method can be used to detect,
 CC monitor progression of, or treat prostate cancers. The antibody may
 CC also be conjugated to a therapeutic agent for use in therapy of prostate
 CC cancers.

XX Sequence 250 BP; 55 A; 75 C; 63 G; 57 T; 0 other;

Query Match 25.9%; Score 236.2; DB 19; Length 250;

Best Local Similarity 98.8%; Pred. No. 6,9e-62;

Matches 238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 413 tgcctcgggtggaagcattatcgttaagacacaggtccagatagact 472

Db 241 TTTTTCGGGTGAGCATGGGACCTTATTCGTTAAGACATCAGGCTCAAGATGAACT 182

QY 473 ttcaagcagaagccttcgcgggagcaagagacagaagaagcgtgagatgagcgtg 532

Db 181 TTGAGCAGAAAGCGCTTCGCGGAGCAAAAGGACGAAAGAGTGAACAGTGCCTGG 122

QY 533 cagcaatcacagccggcgaaggtgtctccgagcctcgatcccccggccggggagctg 592
 |||||||
 Db 121 CAGCAATACAGCCGGGCAAGGTCCTCCGAGCCTCGCATCCCGCGCGGGGCGAGCTG 62
 QY 593 gaggtgctcagaaggtgtcattctgtctccgcaggggcttgaacaacccaagcactca 652
 |||||||
 Db 61 GAGGTGCTCAGAAAGTGCAATTCTGCTTCGACAGGGGCTTGAAACACCAAGGCACCTCA 2
 QY 653 g 653
 |
 Db 1 G 1

RESULT 15
 AAA06412/C
 ID AAA06412 standard; cDNA: 250 BP.
 XX

AC AAA06412;

DT 13-JUN-2000 (first entry)

DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:179.

KM Human; prostate cancer; diagnosis; tumour; gene therapy; detection;

KW immunogenic; cytostatic; vaccine; ss.

OS Homo sapiens.

PN WO200004149-A2.

PD 27-JAN-2000.

PF 14-JUL-1999; 99WO-US15838.

PR 14-JUL-1998; 98US-0115453.

PR 14-JUL-1998; 98US-0116134.

PR 23-SEP-1998; 98US-0159812.

PR 23-SEP-1998; 98US-0159812.

PR 15-JAN-1999; 99US-0232149.

PR 15-JAN-1999; 99US-0232880.

PR 09-APR-1999; 99US-028946.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Harlocker SL, Yugin J, Xu J, Mitcham JL;

PI WPI; 2000-171268/15.

PT New polypeptide useful for treating and diagnosing prostate cancer

PT comprises an immunogenic portion of prostate tumor protein -

PS Claim 50; Page 161; 263pp; English.

XX The present invention describes isolated polypeptides, comprising an
 CC immunogenic portion of a prostate tumour protein (PPP). The polypeptides
 CC and polynucleotides encoding them have cytostatic activity and can be
 CC used in vaccines and in gene therapy. The polypeptides and
 CC polynucleotides encoding them, antigen presenting cells which express
 CC the polypeptides, antibodies against the polypeptides and vaccines
 CC comprising them can be used for inhibiting the development of prostate
 CC cancer in a patient. The polypeptides can be used to generate antibodies
 CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
 CC the polynucleotides encoding the polypeptides can be used as a probe or
 CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
 CC AAY82000 to AAY82020 represent sequences used in the exemplification of
 CC the present invention.

XX Sequence 250 BP; 55 A; 75 C; 63 G; 57 T; 0 other;

Query Match 25.9%; Score 236.2; DB 21; Length 250;

Best Local Similarity 98.8%; Pred. No. 6,9e-62;

	Matches	238:	Conservative	0:	Mismatches	3:	Indels	0:	Gaps	0:
QY	413	tgctctcggtgtgagacatgagaccttattctgttaagaacatcaagctccagatatgaact								472
Db	241	TTTTTTTCGGGTGAGCATGTGGACCTTTATTCTTTAAGACATCAGGCTCCAGATATGAACT								182
QY	473	ttcacagagaagcgctctgcccggagcaaaagagacagaagaagctgagatgaacagtgctg								532
Db	181	TTTCAGCAGAGAGCGCTTGCCGGGAGCAAAAGGAGACAGAAAGTCGAGATGAACAGTGCCTG								122
QY	533	cagcaatcaccaaccgggagcaaggtgtctccgaagcttgcattccccggccggggggcagctg								592
Db	121	CAGCATATCACACCCGGGCAAGGTGTCTCCGACCTTCGCATTCCTCCGGCCGGCGGCAAGCTG								62
QY	593	gaagtgcttcacagaaggtgcatctctgtcttcctcgtcaggggcttgaacaacaaaggcaatcca								652
Db	61	GAGGTGCTCTCAAGAAAGTGCATTCGTGTTCTCTCAAGGCGCTTAAACACCAAGGCAATCCA								2
QY	653	g 653								
Db	1	G 1								

Search completed: March 30, 2002, 13:05:01
Job time: 21298 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 05:39:03 ; Search time 5438.31 Seconds

(without alignments)
1800.082 Million cell updates/sec

Title: US-09-867-034-1

Perfect score: 911

Sequence: 1 ttttttttttgcctgttt.....cacgtgggtccgcggagcgc 911

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

EST.*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estopl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_estcl:*
11: gb_estc2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	568.2	62.4	600	10	AM513333 x043f10.x
2	562	578	10	AM769057	h158b06.x
3	485.6	53.3	500	10	AM769425 h156b06.x
4	482.6	53.0	515	10	AM510389
5	411.4	45.2	421	10	AM007096 x198f12.x
6	408	44.8	430	10	AA633399 np69n05.s
7	402	44.1	402	11	BE002129 7699f01.x
8	401.2	44.0	476	10	AI281282 qk72d10.x
9	396	43.5	461	10	AI143630 qk74c01.x
10	392.2	43.1	463	10	AA315762 EST187535
11	386.8	42.5	476	10	AA552443 nk15d09.s
12	385.4	42.3	390	10	AA314975 EST16849

13	383	42.0	466	10	AM167728	nx48c08.x
14	382.6	42.0	773	11	BG542020	602572132
15	379.2	41.6	460	11	BI114257	602862447
16	377.2	41.4	447	10	AI332337	qp97h12.x
17	371.4	40.8	438	10	AI346752	qp52907.x
18	370	40.6	468	11	BG529874	602558970
19	367.8	40.4	476	10	AA580138	nh77a12.s
20	367.4	40.3	476	10	AA974853	co30g10.s
21	366.8	40.3	457	10	AA716097	z963c06.s
22	364.4	40.0	394	10	AI983974	w020h03.x
23	362.8	39.8	402	10	AI973218	wf53c01.x
24	362.8	39.8	420	10	AI748945	at38h11.x
25	358	39.3	422	10	AI762067	w153b12.x
26	356.4	39.1	420	10	AA507878	h753d09.s
27	348.4	38.2	443	10	BE645173	7664d06.x
28	346.2	38.0	404	11	W60304	z2d9d08.s1
29	345	37.9	405	10	AA552321	nk06h03.s
30	344	37.8	353	10	AI350070	q028b12.x
31	340.2	37.3	412	10	AM517117	xp91f10.x
32	335.4	36.8	454	10	AA953818	co38e03.s
33	334.2	36.7	441	10	AA631397	np86e07.s
34	333.4	36.6	404	10	AA614335	np49c10.s
35	330	36.2	407	10	AA808607	ce56b08.s
36	327.4	35.9	470	10	AA632754	np85d07.s
37	324.2	35.6	482	11	W60395	z2d9d08.r1
38	323.8	35.5	343	10	AI832569	at70b08.x
39	323.2	35.5	433	11	N74131	z275b01.s1
40	320.4	35.2	487	10	AI985964	wf79d08.x
41	317.6	34.9	338	10	AM472805	xq21c03.x
42	317.2	34.8	451	10	AA554791	nk29c08.s
43	317	34.8	323	10	AM291863	ut-h-B12-
44	317	34.8	352	10	AW951538	EST363608
45	314.8	34.6	318	10	AI459674	at84c01.x

ALIGNMENTS

RESULT 1
AM513333
LOCUS
DEFINITION
x043f10.x1 NCI-CGAP Utl Homo sapiens cDNA clone IMAGE:2706763 3'
similar to gb:L15203 INTERSINAL TREPOIL FACTOR PRECURSOR (HUMAN);,
mRNA sequence.

ACCESSION
AM513333
VERSION
AM513333.1 GI:7151411
KEYWORDS
EST.
SOURCE
ORGANISM
Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
1 (bases 1 to 600)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: ggap@fmail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LINL at:
image.llnl.gov/image/html/resources.shtml
Seq primer: -40UP from Gibco
High quality sequence stop: 423.

FEATURES

Source
1..600
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2706763"

[illegible]

JOURNAL COMMENT	Tumor Gene Index
Unpublished (1997)	
Contact: Robert Strausberg, Ph.D.	
Email: c9apb5-r@emil.nih.gov	
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.	
cDNA Library Preparation: Life Technologies, Inc.	
cDNA Library Arrayed by: Greg Lennon, Ph.D.	
DNA Sequencing by: Washington University Genome Sequencing Center	
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LMLL, send email to: info@lml.1ml.gov	
Possible reversed clone: polyT not found	
Seq primer: -400P from Glpco	
High quality sequence stop: 409.	
Location/Qualifiers	
1..578	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:3005363"	
/clone_lib="NCI_CGAP_Col4"	
/tissue_type="moderately-differentiated adenocarcinoma"	
/lab_host="DH10B"	
/note="Organ: colon; Vector: PCMV-SpORt6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 kb. Life Technologies catalog #: 11511-019"	
BASE COUNT	153 a 142 c 142 g 141 t
ORIGIN	
Query Match	61.7%, Score 562, DB 10, Length 578;
Best Local Similarity	98.3%, Pred. No. 3.2e-127;
Matches 568; Conservative	0; Mismatches 10; Indels 0; Gaps 0;
OY 10	tttgctgtgtgtatcatgtttctactgttacaagaacaaaccccggaatgtacaaag 69
Db 1	TTTGGCTTTGGTCAATATGTTACTGTACAAAGAAACAAACCCAGATTAGTACAG 60
OY 70	tattacaagtagcagagagtggtgttgaataaaggaccacttggaaagacgtttatt 129
Db 61	TATTCAACAGTAGCAGAGAGTGGTTGTGAATAAAGACACCTTTGGANAGCACTTTATT 120
OY 130	ggctgtgtctctccacaagaagacttggatttttgaaaactctactgaatgylat 189
Db 121	GGCTGTGCTGTCTTCACCAAGAAAGACTGTGATTTTTGAATACTTACCTGAATGTAT 180
OY 190	ttttctgtcttcccgaggaagcggaacttaccgtgttctcaggcttctcgtgaagtgg 249
Db 181	TTTTTCTGTCTTCCCGAGGAAGCGGCACTTACAGTGTCTTCAAGGCTTTCCGTACAGTGG 240
OY 250	gtgcagactgtgatccaatatctctgtcagactgacgtctccttgaaggagcttctcc 309
Db 241	GTGCAGACTGTGATTCAAAATATCTCTTGACATGCATGCAGTCACTCTTAGAGGAGTCTTTTCC 300
OY 310	tgcccttgaagccttggcagactctccctctgacacacctcccgccctctccacgacgag 369
Db 301	TGCCCTTGAGGCTTGAGCAGACTCTCCCTTGACACCTCCGCGCTCTCCACAGACGAG 360
OY 370	cagaaataaagacacaaactctgaagaagctcaggcagcagaagaactgtctcgtgttgaagca 429
Db 361	CAGAAATTAACCAACAACCTCAGAAAGTCTCAGGACGACGAAAGTCTCTCGGTGGAGCA 420
OY 430	tggaaccttattcgttaagaacatcagagctccagatatagaacttcaagcagaagcgcttg 489
Db 421	TGGGACCTTTATACGTTTAGACATCAGGCTCTCAATATGAACTTTCAGCAAAACGGCTTG 480
OY 490	ccggagcaagggagcagaagaagcttgatataaagcttgcttgcgacgaatcaacgcccgg 549
Db 481	CCGGGACGAAGGAGCAGAGAGAACTGAGATTAACAAGTGTCCCTGGCAGCAATTCACAGGCGGG 540
OY 550	caagagtgtctcgaagcctcgcatcccccgcgacgggggc 587

Db 541 CAAGATGCTTGAAGCTTCGATCCGACCGCCAGGAGGC 578

RESULT 3
LOCUS AW769425 500 bp mRNA EST 04-MAY-2000
DEFINITION h156b06.x1 NCI_CGAP Col4 Homo sapiens cDNA clone IMAGE:3005171 3' similar to gb:LI5203 INTERINAL TREFOIL FACTOR PRECURSOR (HUMAN); mRNA sequence.
ACCESSION AW769425 GI:7701456
VERSION AW769425.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 500)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -400P from GIBCO
High quality sequence stop: 377.

FEATURES
Source location/Qualifiers
1..500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3005171"
/clone_lib="NCI CGAP Col4"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"

BASE COUNT 133 a 117 c 116 g 134 t
ORIGIN

Query Match 53.3%; Score 485.6; DB 10; Length 500;
Best Local Similarity 98.2%; Pred. No. 1.6e-108;
Matches 491; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 10 ttgccttgcttgctcataatgttctactgtacaaagaacaaacccaggaatagatcaag 69
Db 1 TTTGGCTGTTTGTATATGTTTACTGTACAAAGAAACAAACCCAGGAATATGACAG 60

QY 70 tatggaagaagtagcagaggtgtgttgaataaaggaccacttggagaagacagtttatt 129
Db 61 TATTGAAGAGTAGAGCAGAGTGGTGTGAATTAAGACCACTTGGAAAGACAGTTTAT 120

QY 130 ggcctgctcttccacaaagaagactgtgatttttgaanaactctacctaataatgcat 189
Db 121 GCGTTGCTGTCTTCACCAAGAAAGACTGTGATTTTGAACAACTTCTACCTGAATGTAT 180

QY 190 tttctgcttcccgaggaagcggcacttaacagtgtctcttagagcttctctgtgacgtg 249
Db 181 TTTTCTGCTTCCCGAGGAAGCGGCACTTACAGTGTCTTACGCTTCTTACGACGTG 240

QY 250 gtgcagagctgattcaaatatattctctgcatgcaactgcaagctcccttaaggagctttcc 309
Db 241 GTGCCAGTGTGATTCAAAATATCTTCGATGCACTGACGCTCTTAGGAGGATCTTTCC 300

QY 310 tgccttagagccttgagcagactccctctgacacccctccgcgcctctccacagacgag 369
Db 301 TGGCCTTAGGGCTGTGGAGACTCTCCCTGACACCCCTCCGCTCTCCACAGACGAG 360

QY 370 cagaataaagacacaacctcagaaagltcaggaacgaagaactgtctcgtgtgagca 429
Db 361 CAGAAATTAAGCACACACACTCAGAAAGTCTCAGGACGACGAAGAACTGCTCGGTGAGCA 420

QY 430 tggagccttattcgtttaagacatcaggtcccaagatagacttccagcaagaagcctg 489
Db 421 TGGACCTTTATTCGTAGACATCAGCTTCATATGAACTTTTACGAGAAGCGCTTG 480

QY 490 ccgaggaacaaaggagacagaa 509
Db 481 CCGGAGCAAGAGGACAGAA 500

RESULT 4
LOCUS AW510389 515 bp mRNA EST 03-MAR-2000
DEFINITION x198f12.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2682767 3' similar to gb:LI5203 INTERINAL TREFOIL FACTOR PRECURSOR (HUMAN); mRNA sequence.
ACCESSION AW510389
VERSION AW510389 GI:7148467
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 515)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/lresources.shtml
Seq primer: -400P from GIBCO
High quality sequence stop: 310.

FEATURES
Source location/Qualifiers
1..515
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2682767"
/clone_lib="NCI CGAP Ut1"
/tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

BASE COUNT 139 a 119 c 116 g 141 t
ORIGIN

Query Match 53.0%; Score 482.6; DB 10; Length 515;
Best Local Similarity 98.0%; Pred. No. 8.8e-108;
Matches 499; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 16 tgtgttcataatgttctactgtacaaagaacaaacccgggaatagatcaagratga 75
Db 4 TTTTGTTCATATATGTTTACTGTACAAAGAAACAAACCCAGGAATATGATTTGA 63

QY 76 acagtagcagagtggtgttgtaataaagaccacttggagaagacagtttattgtcttg 135

|||||
Db ACATGACGAGAGTGTGTAATAAGACACTTTGGAAGACAGTTTATTGGCTTG 123
136 ctgtctccacaagaagacttgatcttttgaactctactgaatgatcttc 195
|||||
Db CTGCTTCACCAAGAAAGACTGTGATTTTGAAGAACTTACTGAAATGATTTTTC 183
124
196 tgcctccaggaagacgagactacagtgctccaggtcttcctgagcgtgggcca 255
184 TGTCTTCCAGAGGAGGAGCACTTACAGTGTCTTCTAGGCTTCTTGAGCGTGGGCCA 243
QY 256 gtcgtggttcaaatatccctgtcagtcagtcagctccttaggaagtccttcctgcct 315
Db 244 GTCGTGATCAAAATATCTTGACATGACATGACACTCTTAAAGGAGTCTTTCTGCTCT 303
QY 316 tgaagctggcagacactccctcagacacccctccctccctcccaagacagagaaa 375
Db 304 TGAGGCTGGGCAAGACTCTCCCTGACACCTCCGCGCTTCTCAGCAGCAGCAAAA 363
QY 376 taaagcaaacctcagaagatctcaggaagaaactgtccctcgggtggagcaatggagac 435
Db 364 TAAAGCAACACTCAGAGACTCATGACAGAAAGACTGTCTCGGGTGAGCATGGGAC 423
QY 436 ctattatcgttaagactcaggtccagatagactcttcagcagaagcgttgcggga 495
Db 424 CTTTATTCTGTTAAGACATCAGGCTCAGATATGAACCTTTAAGCAGAAAGCGCTTGCGGA 483
QY 496 gcaaaaggacagaagaagctgagatgaaca 524
Db 484 GCCAA-GGACAGAAAGCTGAGATTAAAA 511

RESULT 5
AM007096 421 bp mRNA EST 10-SEP-1999
LOCUS wt09907.x1 NCL_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2506933 3',
DEFINITION mRNA sequence.
ACCESSION AM007096
VERSION AM007096.1 GI:5855874
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 421)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE Tumor Gene Index
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. 421

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2506933"
/clone_lib="NCL_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT7AD-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was

ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library went through one round of
normalization.
BASE COUNT 108 a 99 c 94 g 120 t
ORIGIN

Query Match 45.2%; Score 411.4; DB 10; Length 421;
Best Local Similarity 98.6%; Pred. No. 2.4e-90;
Matches 415; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 11 tgcgtgttgcataaagtattactgtacaaagaacaaaccaggaaatgacaagt 70
Db 1 TTTTGTGTTGTCATATATGTTACTTACTACAAAGAAACGAACTCAGGATGATACAGT 60
QY 71 attgaacagtagagagtggtgtgtaataaaggaccacttggagagagtttatg 130
Db 61 ATTGAACAGTAGCAGAGGTGTTGTAATTAAGACCACTTTGGAAGACAGTTTATTTG 120
QY 131 gcttgcgtctcacaagaagaactgtgatcttgaagaactctacgtgaatgtatc 190
Db 121 GCTTGTGCTCTTCAACCAAGAAAGACTGTGATTTTGAACCTTCACTCAAGATGATTT 180
QY 191 tttctgtcttccaggaagagcgacactacagtgctcctaagcttcctgtgagtg 250
Db 181 TTTTCTGCTTCCCGAGGAAGCGGCACTTACAGTGTCTTCTTCTTCTGACGTGGG 240
QY 251 tgcagctgtgattcaaatatctctgcatgcaactgagctccttggaggtcttctc 310
Db 241 TGCCAGTCTGATTCAAATATCTTGCATGACACTGACGCTCTTGAAGAGTCTTTTCT 300
QY 311 gcccttaagccttggcagacactcccttacaacccctccctcccaagcagcagc 370
Db 301 GCCTTGAGGCTTGGGCAAGACTCTCCCTGACACCTCCGCGCTTCCACAGCAGGAGC 360
QY 371 agaaataaagcacaaacctcagaagatctcaggaagaagaactgtcctgggtggagcat 430
Db 361 AGAATTAACACCAACCTCAGAAAGTCTCAGCAGCAAGAAAGCTCTCGGTGGAGACAT 420
QY 431 g 431
Db 421 G 421

RESULT 6
AA633399 430 bp mRNA EST 28-OCT-1997
LOCUS np69H05.s1 NCL_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1131609 3',
DEFINITION similar to gb:L15203 INTERSTINAL TREFOLIL FACTOR PRECURSOR (HUMAN);,
mRNA sequence.
ACCESSION AA633399
VERSION AA633399.1 GI:2555259
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 430)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE Tumor Gene Index
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Insert Length: 563 Std Error: 0.00
Seq primer: -40ml3 fwd. E7 from Amersham
High quality sequence stop: 423.
Location/Qualifiers

FEATURES

SOURCE

1. 430
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1131609"
/sex="female, pooled"
/issue_type="breast"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from pooled bulk
breast tumor tissue, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. This library is the normalized version of
NCI-CGAP_Brl.1. Library was constructed by Bento Soares
and M. Fatima Bonaldo."

BASE COUNT 94 a 125 c 136 g 75 t
ORIGIN

Query Match 44.8%; Score 408; DB 10; Length 430;
Best Local Similarity 99.5%; Pred. No. 1.6e-89;
Matches 430; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 419 cgggtgagcatgggacattatctgttaagacalcagctccagatatagaatttcagc 478
DB 1 CGGGGGAGCATGGGACCTTATTCTTAAAGATCATCAGCTCCAGATATGAACCTTTCAGC 60
QY 479 agaagcgttcgaggaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 538
DB 61 AGAAGCGCTTCCGGGAGCAAGAGGAGCAAGAAAGCTGAGATGACAGTCCCTGGCAGCAA 120
QY 539 tcacagccgggcaagggtgctcgcagctcgcacatcccccgcgcgggggagctgaggtg 598
DB 121 TCACAGCGGGGCAAGGAGGTCCTCCAGGCTCCGATGCCCGGCGC -GGGGCAGCTGAGGTG 179
QY 599 cctcagaagtgatctcgtctcctcgcaggggcttgaacaacccaaggaactccaggaatc 658
DB 180 CCTCAGAGGTGATCTGCTTCTCCGACGGGCTTGAACCAAGGCACTCCAGGAGATC 239
QY 659 ctgagatcaagaagcagcccggtgtgtgactcctcgtgggggtgacatgaggggtgagcgg 718
DB 240 CTGAGATCAAGACACACACCCCGGTTGTGCACTCTTGGGGGTGACATGGGGGTAGCCG 299
QY 719 cagtcacccctgtcctgtcgtcgcagcagcagcactgttgcagctgtccagacaagaagcc 778
DB 300 CAGTCCACCTGTCCTGTGGCTGGCAGGCGACACCTGTTGCACTGTCACAGCAAGAGCC 359
QY 779 ctgtcagctccagaagccttgcgtgagcagggccagactactcctcagcagagctgag 838
DB 360 CTGTGAGCTGGCAGAGCCCTTGTCTGGAGACAGGCCACCTAC -TCTCAGACAGAGCTGAG 418
QY 839 gacagcaagggc 850
DB 419 GACAGCAAGGCC 430

RESULT 7
LOCUS BF002129 402 bp mRNA EST 06-OCT-2000
DEFINITION 7g99101.x1 NCI-CGAP_Col16 Homo sapiens cDNA clone IMAGE:3314617
ACCESSION BF002129
VERSION BF002129.1 GI:10702404
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLT, send email to:
info@image.lnlt.gov
Seq primer: -40UP from Glibco.

FEATURES

SOURCE

1. 402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3314617"
/clone_1lb="NCI-CGAP_Col16"
/issue_type="Colon tumor, RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP_Col16 was
prepared, and ss circles were made in vitro. Following NRP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids: 1057416-1061255, and 1144584-1145551).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 107 a 97 c 87 g 111 t
ORIGIN

Query Match 44.1%; Score 402; DB 11; Length 402;
Best Local Similarity 100.0%; Pred. No. 4.8e-88;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ttgctgtctgttcataatgttactgtacaaagaacaaacccaggaatagatcaag 69
DB 1 TTTGGCTGTTGTCATATGTTTACTGTACAAAGAAACCAACCCAGATATGATACAG 60
QY 70 tatgaacagtaagagagatgtgtgtaataaaggaccacttggagaagacttatt 129
DB 61 TATTGAAACAGTAGCAGAGGTGTTGTGAATAAAGACCACTTGGAGACAGTTTATT 120
QY 130 gactgtctgtcttcacaaagaagcctgtgatttttgaactctacactgaatgat 189
DB 121 GGCCTTGCTGCTTCCACAGAAAGACTTGTGATTTTAAANACTTCTACCTGAATGTAT 180
QY 190 ttctctgtcttcacaaagaagcagcacttacagtgctcctcaggtcttcgtgagctgg 249
DB 181 TTTTCTGCTTCCCGAGGAAGCGGCACTTACAGTGTTCCTAGAGCTTCCTGATCGTGG 240
QY 250 gtgcagctctgattcaaatatctctgcacatgactgacagctcctctaggaagctcttc 309
DB 241 GTGCCAGTCTGATTCAAATAATCTTCATGACAGCTCCTTATAGGAGTCTTTTCC 300
QY 310 tgccttgaagccttggagaagcctcctcctgaacccctccgcctctccagaagcagcag 369
DB 301 TGCCCTTGAAGGCTGGGAGACTCTCCCTGACACCCCTCCGCCCTCTCCACGAGCGAG 360
QY 370 cagaataaagcacacacctcagaagcttcagggcaggaaga 411
DB 361 CAGAAATTAAGCACACACTCAGAAAGTCTCAGGACGAGCAAGAA 402

RESULT 8
A1281282

LOCUS AI281282 476 bp mRNA EST 23-NOV-1998
DEFINITION gb72d10.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1874515 3'
similar to gb:115203 INTERSTITIAL TREFOIL FACTOR PRECURSOR (HUMAN);,
mRNA sequence.
ACCESSION AI281282
VERSION AI281282.1 GI:3919515
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 476)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbdp/imagenet/image.html
Seq primer: -400p from Gibco
High quality sequence stop: 423.
Location/Qualifiers
1..476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1874515"
/clone_lib="NCI_CGAP_Co8"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon. Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 106 a 138 c 148 g 84 t
ORIGIN
Query Match 44.0%; Score 401.2; DB 10; Length 476;
Best Local Similarity 90.7%; Pred. No. 7.5e-88;
Matches 468; Conservative 0; Mismatches 3; Indels 45; Gaps 2;
396 tctcagcagcaagaactgtccctcgtgagcagctggaacttatcgttaagactca 455
|||||
Db 6 TTTCAGGACGCAAAATGTCCTGGGTGGAGCATGGACTTATTCGTTAAGACATCA 65
456 gqctccagatatagaacttcaagcagaagcgctgcgagagcaaaaggacagaactg 515
|||||
Db 66 GCGTCCAGATATGAACCTTTCAGCAGAAAGCGCTGCCGGAGCAAAAGGACGAAAGCTG 125
516 agatgaacagtgcctgcagcaaatcacagccgggcaagggtgctcgagctcgatccc 575
|||||
Db 126 AGATGAACACTGCTGCGCAGCAATCAACGCGGGCAAGGGGTGCTCCAGCTCCATCCC 185
576 ccggccgggggagcagctggaagtgctcagaaggtgcatctcgtctcaggggtga 635
|||||
Db 186 CCGGCGGGGAGCTGAGAGTGCCTCAGAAAGTGCAATTCCTGCTGCAAGGGCTTGA 245
636 aacaccaagcactccagggatcctgagatcaaaagcagcagcccggtcttggaactcct 695
|||||
Db 246 AACACCAAGGACATCCAGGAGTCTCGAGTCAAAGCAGCAGCCCGGTTGTGACATCCT 305
696 tgggggtgacatggggtagccgagtcacaccctgtcttgctgagcaggaactggt 755

|||||
Db 306 TGGGGGTGACATGGGGGTGACCCGAGTCACCCCTGCTTGGCTGGACGCGACACATGCT 365
QY 756 ttgagctgtcccaagaagaccctgtcagctgccagaagccctgtcgtgggaagggccac 815
|||||
Db 366 TTGCA-----GACAGGCCAC 361
QY 816 gtacttcacagcagagctgtagagacaggaagccagaccagcccaatgcaagcg 875
|||||
Db 382 GTAC-TCCCTCAGCAGACTGTGAGGACAGCAAGGCGCAGACAGCCCGCATGCAAAAGCG 440
QY 876 ctctgcaagccatgacacccgtggtcgcggagcgc 911
|||||
Db 441 CTCGCGCAGCATGACCAACCGCTGGCTCCGCGAGCG 476

RESULT 9
AI143630 461 bp mRNA EST 26-OCT-1998
LOCUS AI143630
DEFINITION gb74c01.x1 Soares_fetal_heart_NBH19W Homo sapiens cDNA clone
IMAGE:1705824 3' similar to gb:115203 INTERSTITIAL TREFOIL FACTOR
PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION AI143630
VERSION AI143630.1 GI:3665439
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 461)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 510 Std Error: 0.00
Seq primer: -40m3 fwd. ET from AmerSham
High quality sequence stop: 416.
Location/Qualifiers
1..461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1705824"
/clone_lib="Soares_fetal_heart_NBH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart. Vector: pT73D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTTCACATCTGAAGTGGAGCGCGCCGATCTTTTATTTTATTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M. Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBH19W."

BASE COUNT 103 a 136 c 146 g 76 t
ORIGIN
Query Match 43.5%; Score 396; DB 10; Length 461;
Best Local Similarity 91.1%; Pred. No. 1.4e-86;
Matches 461; Conservative 0; Mismatches 0; Indels 45; Gaps 2;
399 caggcaggaagaactgtccctcgtgtagagatggagacttatcgttaagacatcagc 458
|||||
Db 1 CAGGACGAAGAACTGTCCTGGGTGGAGCATGGACTTATTCGTTAAGACATCAGCG 50

QY	459	tcacgaatgaacttcacgaagaagccttgcgagagcaaaaggaacgaaaagctgaga	518
Db	61	TCcGaTATgAcTCTTAgcAGaAGAGcCTTgcGgAGcAAAGGAGcAGAAAGcTGaA	120
QY	519	tgaacatgacctgacgaatcacaagccgagcaagagtgtccgaagcctgcaccccg	578
Db	121	TGAACAGTcCTGcGACCAATCAcAGcCGGgGAAGGATGCTCCGAGCTGCATCCCGG	180
QY	579	gcgcggggcagctggaagtgcctcaagaagtgatcttgccttcgaaggcttgaac	638
Db	181	GCcGgGGGcAcTcGAGcTGcCTTAgAGgTGcATTTGcCTTcGAGGgGcTTAAAc	240
QY	639	accaagagcactccagggatccctgagtcacaaagcagacgcccgcgattgtacctcttg	698
Db	241	ACCAAGGcAcTCCAGGgATcCTGAGTCAAAGcAGcAGcCCGgTGTTCATcCTTGG	300
QY	699	gggtgacatggggttagcccgacgtccaccctgtccttgcttgacagcgacactglttg	758
Db	301	GGGtGAcATGGGGGTAGCCGcAGTCcACCTGTGTGGCTGGcAGGcGCACATGGTTTG	360
QY	759	cagctgtccacagaacaagccctgtcagcttgcaagagccctgtgtggaaagggccacgta	818
Db	361	CA-----GACAGGcCCAcGTA	376
QY	819	cttcctcagaagaagcttgaagagacaagaagcccaagaccccaagatgcagaagcctc	878
Db	377	C-TcCTcAGAGAcTcGAGAGAGcAGCAAGcCCAGAcACcCCcAGATGcAGAGcGCTC	435
QY	879	tggcagcagatgacaacgcgttggtccc	904
Db	436	TGGcAGcCATgAcCCAcCTGcGGCTCC	461

RESULT	10	AA315762/c	LOCUS	AA315762	463 bp	mRNA	EST	19-APR-1997
DEFINITION				EST187535	Colon carcinoma (HCC) cell line II Homo sapiens	CDNA	5	
				end similar to similar to trefol factor, intestinal, mRNA sequence.				

ACCESSION	AA315762
VERSION	AA315762.1
KEYWORDS	EST.
SOURCE	human.

ORGANISM	REFERENCE
Homo sapiens	Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi	1 (bases 1 to 463)
Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.	

C.J., Lee, N.H., Kirkness, E.F., Weinstein, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Whai, C., Clifton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodde, A., Grehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaro, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudak, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Benavrik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferlie, A., Fisher, C., Hastings, G.A., He, M.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

TITLE Initial assessment of human gene diversity and expression patterns
JOURNAL Nature 377 (6547 Suppl.), 3-174 (1995)
MEDLINE 96026280
COMMENT Other_ESTS: THC178105
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavetigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers

```

Source
1. .463
/organism="Homo sapiens"
/db_xref="ATCC (host):111494"
/db_xref="taxon:9606"
/clone_lib="Colon carcinoma (HCC) cell line II"
/tissue_type="colon"
/cell_type="KM12C"
/cell_line="KM12C(HCC)-parental human colon carcinoma
;Dukes B2"
/note="Organ: colon; Vector: pbluescript SK-; Site_1
EcorI; Site_2; XhoI"
BASE COUNT      77 a      146 c      134 g      105 t      1 others
ORIGIN

```

Query Match	43.1%	Score 392.2;	DB 10;	Length 463;
Best Local Similarity	90.4%	Pred. No. 1.2e-85;		
Matches 459; Conservative	0;	Mismatches 4;	Indels 45;	Gaps 2;

Oy	399	caggcagaaagaacgctccctcggtgtagacatggaacattatccgttaagaacatcaagc	458
Db	463	CAGGCACCAAGAAATGTCTCGGGGAGGCAATGGGACCTTTATTCGTTAAACATCAGGC	400
Oy	459	tccagatatgaacttccagcagaagcgcttgcgcggagcaaaagygacagaaaagcttgaga	518
Db	403	TCCACATATTGAACTTTTCAGCAGAAAGCGCTTCCGGGACCAAGGGACAGAAAGCTGAGA	344
Oy	519	tgaacagagccctggcagagaatcacagccgggcagaaggtggtccagacctgcatcccccg	578
Db	343	TGAACAGTGCTCGCAGAGCAATTCACAGCCGGGCAAGGGTGTCTCCGAGCTTCCATGCCCG	288
Oy	579	gcgcggggcagccttgagagtgtccctcagaaggtgtcatctgtcttctcgcagggcttgaac	638
Db	283	GCCTGGGGCAGCTGTGAGGTGCTTCAGAGAGTGATTCGTCTTCGCGAGGGGCTTGAAC	224
Oy	639	accaaggaaccccaaggagatctctgaagtcaaaagcagccccgggttgttgaacctcttg	699
Db	223	ACCAAGGCACTCCAGGGATCTCTGGAGTAAAGACAGACCCCGGATGTTGCACTCTTGG	164
Oy	699	gggtgacatlggggttagccgcagttccacccttcttgcgtgcagcaggcacactggttg	758
Db	163	GGGTGACATGGGGGAGCGCCAGTCCACCTCTGTCCTTGCTGGCAGGCACTAGTGTG	104
Oy	759	cagctgtcccaagacaagccctgttcagctgtccagagcccttgcctgggacaaagcccaagta	818
Db	103	CA-----GACAGGCCACACTA	88
Oy	819	cttccctcagcagagcttgtagagacagcaagcaccagcccccagcttgcagagcgctc	878
Db	87	C-TCTCTAGCAGAGCTGTGAGGACAGCAGGCGCAGGACCCAGCATGCAGAAAGCTC	29
Oy	879	tggcagcgcatgacccacgcttggtctccgg	906
Db	28	TGGCAGCCATGACACCGCTGGGACTTCGG	1

RESULT	11
AA552443	
LOCUS	
DEFINITION	n115d09.s1 NC1_CGAP_CO2 Homo sapiens CDNA clone IMAGE:1013585 3' similar to gb:U15203 INTERSTITIAL TGFBI-FIL FACTOR PRECURSOR (HUMAN); mRNA sequence.
ACCESSION	AA552443
VERSION	AA552443.1 GI:2322697

	KEYWORDS	EST.
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
COMMENT	Tumor Gene Index Unpublished (1997)	
	Contact: Robert Strausberg, Ph.D. Email: cga@db-rt@mail.nih.gov Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.	
	CGNA Library Preparation: Stratagene, Inc., David B. Kitzman, Ph.D.	
	CGNA Library Arraying: Greg Lennon, Ph.D. DNA sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LML at: www-bio.lml.gov/bhrp/image/image.html Insert Length: 666 Std Error: 0.00 Seq primer: ~40mJ3 fwd. ET from Amersham High quality sequence stop: 321.	
FEATURES	Location/Qualifiers	
source	1..476 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1013585" /clone_1id="NCI.CGAP_Co2" /clone_1ld="NCI.CGAP_Co2" /tissue_type="tumor" /lab_host="SOBR (kanamycin resistant)" /note="Organ: colon; Vector: Bluescript SK-; Site.1: EcoRI ; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Bulk colon villous adenoma. 5' adaptor sequence: 5' GAATTCGGCACAGAC 3' 3' adaptor sequence: 5' CTCAGAATTGTGTTTTTTT 3' Average insert size: 1.1 Kb."	
BASE COUNT	107 a 136 c 152 g 81 t	
ORIGIN		
Query Match	42.5%; Score 386.8; DB 10; Length 476; Best Local Similarity 87.8%; Pred. No. 2.5e+84; Matches 453; Conservative 0; Mismatches 17; Indels 46; Gaps 1;	
OY	366 tctatgacgcacaagaactcgtcttcggcgaggagcataggaccttatcttgtaagcacatca 455	
DG	1 TTTTAGGCACAAGAAGACTTGCTCCGCCGGTGAGCAACAGGACCCTTAATGCTTAAGACATCA 60	
OY	456 ggctccagatatgaactttcagcaaaagcgtctgccggagagcaaaaaggagcagaagtcty 515	
DG	61 GGCTCCACATATGAAACTTTCACGACAAAACCGCTGCCGGAGAGCAAAGGAGCAGAAAAAAGCTG 120	
OY	516 agatgaacagtgccttgccagcaaatcaaacagccggagcaaggtgtgctccgagcctcgatccc 575	
DG	121 AGATGAACAGTGCCTCGCGACGACATACACAGCCGGCAAGGGTGTCTCCGAGCCCTGCATCCC 180	
OY	576 ccggccggggggcagcgttgttaggtgctcccaaaggtgatcttcgctcccgcaagggtcttya 635	
DG	181 CCGGCCGGGGGCAAGGCTGGAGGGTGCTCTAGAAAGTGCATCTTGCTTCCTGCAAGGGGCTTGA 240	
OY	636 aaccaccaagcactccaaggaatccctcyaggtcaaaacagcagcccggtgtgtgcaactcct 695	
DG	241 AACACCAAGGCACTCCAGAGATCCCTGGAGTCAAACACAGACAGCCCGGTTGTCGACTCTCT 300	
OY	696 ttgggggtgacatggtggttagccgacagtcacaccgtgtccttggtctygctygcaagcagtgt 755	
DG	301 TTGGGGGTGACATGGGGGTGAGCGCGCATGCACCTCTGTTGGCTGGCAGCGCACACTGCT 360	
OY	756 ttgcaagctgtccccagacaagaaccctgttaagctgcagagagcccttgtctygaaggyccaac 815	
DG	361 TTTCGAGACAGGCGCACTG----- 378	

QY	816	gtactctcagcagcagctgtgagagcagcagcagcagccagcagctgtcagcagc	875
Db	379	-----CTCTCTACGACAGCTGTGAGACACACAGCCAGGACCAAGCATGACAGCG	434
QY	876	ctctgcagccatgagcaccgcgtggtcgcgggagc	911
Db	435	CTCTGAGAGCCATGACCATCGGGCTCTCGGAGCGC	470
RESULT	12		
LOCUS	AA314975/c		
DEFINITION	AA314975	390 bp mRNA	EST 19-APR-1997
VERSION	AA314975	ESR186849 HCC cell line (metastasis to liver in mouse) II Homo sapiens cDNA 5' end similar to similar to trefol1 factor,	
KEYWORDS	AA314975.1	GI:1967374	
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Maniayota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 390)		
	Adams M.D., Kerlavage A.R., Fleischmann, R.D., Fuldner, R.A., Bult C.J., Lee N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, C.O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wha, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodok, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmarois, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudak, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dinke, D., Feng, D.-F., Ferlie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, T., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.		
TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence		
JOURNAL	Nature 377 (6547 suppl), 3-174 (1995)		
MEDLINE	96026280		
COMMENT	Other_ESTs: THG178105 Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: akerlavet@tigr.org For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tcdb/hgi/hgi.html) Seq primer: M13 Reverse.		
FEATURES	Location/Qualifiers		
SOURCE	1..390 /organism="Homo sapiens" /db_xref="ATCC (ihost):113179" /db_xref="taxon:9606" /clone_lib="HCC cell line (metastasis to liver in mouse) II" /tissue_type="colon" /cell_type="KM12SM" /cell_line="KM12C(HCC)metastasis into mouse (liver)" /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	71 a 108 c 113 g 95 t 3 others		
ORIGIN			
Query Match	42.3%	Score 385.4;	DB 10; Length 390;

Sat Mar 30 14:59:28 2002

us-09-867-034-1.rst

Page 11

Search completed: March 30, 2002, 09:31:07
Job time: 13924 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 09:32:17 ; Search time 5438.31 Seconds

(without alignments)
5157.205 Million cell updates/sec

Title:

US-09-867-034-5

Perfect score:

2610

Sequence: 1 gatgtggcagcctcagag.....ccgagatgctcaggagagt 2610

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*

1: em_estfun:*

2: em_esthum:*

3: em_estlin:*

4: em_estom:*

5: em_estpl:*

6: em_estda:*

7: em_estro:*

8: em_estov:*

9: em_hlc:*

10: gD_estl:*

11: gD_est2:*

12: gD_hlc:*

13: gD_gss:*

14: em_gss_fun:*

15: em_gss_hum:*

16: em_gss_lin:*

17: em_gss_pln:*

18: em_gss_pro:*

19: em_gss_rtd:*

20: em_gss_vrt:*

21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	653.6	25.0	877	11	BE872086 601448177
2	555.4	21.3	952	11	BC386415 602455641
3	433.7	16.7	621	10	AW969591 EST381668
4	433.6	16.6	666	10	AI732648 nm73f01.x
5	368	14.1	536	10	AI801356 t088c05.x
6	358.6	13.7	525	10	AI222799 qp39d10.x
7	351	13.4	450	10	AW820993 RC2-ST030
8	343.4	13.2	1779	12	AK007445 M8 muscu
9	325	12.5	429	10	AI821412 nes8d09.x
10	320.6	12.3	517	10	AW610230 RC2-ST030
11	313.4	12.0	449	11	BF431333 naa43f01
12	312	12.0	312	10	AA328866 EST32671

C	13	308.2	11.8	418	10	AI281212	AK58609.x
C	14	306	11.7	491	10	AI791944	AI791944 nm73f01.y
C	15	305	11.7	450	10	AA527898	AA527898 nes8d09.s
C	16	301	11.5	546	10	AW140090	AW140090 UI-H-Btl-1
C	17	280.6	10.8	470	10	AA576913	AA576913 nm73f01.s
C	18	256.4	9.8	374	11	W87535	W87535 zh55g04.s1
C	19	248.4	9.5	427	10	AI394677	AI394677 t924p01.x
C	20	225.6	8.6	443	11	BF000690	BF000690 7h34p01.x
C	21	192.8	7.4	226	10	AI739162	AI739162 w127a01.x
C	22	183.8	7.0	236	10	AA583007	AA583007 nt76f03.s
C	23	170.4	6.5	2237	12	AK011082	AK011082 Mus muscu
C	24	167.4	6.4	199	10	AW085900	AW085900 xc60f12.x
C	25	149.2	5.7	257	11	BF766225	BF766225 IL2-CS005
C	26	149.2	5.7	257	11	BF766229	BF766229 IL2-CS005
C	27	148.4	5.7	644	11	BF579351	BF579351 602095640
C	28	124.6	4.8	441	11	H80869	H80869 yu58h04.s1
C	29	115	4.4	707	10	AW390130	AW390130 CM4-ST017
C	30	108.2	4.1	515	10	BE232772	BE232772 137905 MA
C	31	106.8	4.1	191	10	AW376331	AW376331 RC6-CM020
C	32	92.8	3.6	569	10	BE235607	BE235607 143157 MA
C	33	88.6	3.4	960	11	BE788920	BE788920 601585470
C	34	86.8	3.3	883	11	BI257633	BI257633 602967550
C	35	83	3.2	822	11	BE896215	BE896215 601438928
C	36	81.4	3.1	631	10	BE280813	BE280813 601155253
C	37	81.4	3.1	682	11	BF196339	BF196339 7n68f01.x
C	38	81.4	3.1	873	11	BF984445	BF984445 602307990
C	39	79.8	3.1	744	11	BG327067	BG327067 602426275
C	40	79.6	3.0	941	11	BG341123	BG341123 602462772
C	41	78.8	3.0	561	11	BG229942	BG229942 mac24f09.
C	42	78.8	3.0	665	13	AZ305382	AZ305382 IM0006J04
C	43	78.2	3.0	504	11	BE901092	BE901092 601674449
C	44	77.2	3.0	376	10	AI578443	AI578443 UI-R-Y0-U
C	45	77.2	3.0	400	10	AI577576	AI577576 UI-R-Y0-U

ALIGNMENTS

RESULT 1
BE872086 877 bp mRNA EST 20-OCT-2000
601448177F1 NIH_MGC_65 Homo sapiens CDNA clone IMAGE:3852018 5',
LOCUS
DEFINITION
mRNA sequence.
ACCESSION BE872086
VERSION BE872086.1 GI:10320862
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 877)
NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
Contact: Robert Strausberg, Ph.D.
COMMENT
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: L1AM9573 row: 1 column: 19
High quality sequence stop: 645.
Location/Qualifiers
1..877
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3852018"
/clone_id="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"

FEATURES

source

```

/note="Organ: colon; Vector: pcMV-SpOrt6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 1.8 Kb. Library constructed by Life
Technologies."
BASE COUNT      180 a      261 c      300 g      136 t
ORIGIN

```

Query Match	25.0%;	Score 653.6;	DB 11;	Length 877;
Best Local Similarity	95.1%;	Pred. No. 9.3e-126;		
Matches 761;	Conservative	0;	Mismatches 29;	Indels 10;
				Gaps 8;

	LOCUS	DEFINITION	RESULT 2	EST	IMAGE
OY	1190	ggaagagaagatgacagacagacagatgaagcttcaagaatgacagatgcaagacaga	1245		
Db	33	GGAGAGAAAGTGTACGACGACAGCAAGTAAGCTTCTACAAAGATGTACAGATCA--CCAGA	91		
OY	1250	gcagagtgtgaatgaagaactttgttagctcctgaagacatttctccctgaactgtltcaag	1309		
Db	92	GCAGGTGTGAAGATGAAGACTTTTGTGAGTCTGAGGCCATTTCT--CTGACCTTGTCAAG	150		
OY	1310	aagatccggtatltccaacagaaatactaccctcccaagatgatavagaaatgtltca	1369		
Db	151	AAGATCCGTATTTCACACAGAAATAATCAACCCCTCCACAGATATGAGATGTTCTC-	209		
OY	1370	agaaacttggcgatcatcatctggaataatgaattcaaggggtcaactctgagccctcaag	1429		
Db	210	AGAAATCTTGCGCATCATCTGTGAAATTAATTCAGGGGTATCATCTGTGAGCCCTTCAGAC	269		
OY	1430	cgcacagccgagacacttgatctctcgaagaagaagaaagtcagtgaagtacacccgcaga	1489		
Db	270	CGCCAGCCGGAG--CTGSGTTCTCTCGGAAGACAGGAAGTCAAGTAGATACACCGGGAGA	328		
OY	1490	agaaagccttgcaagacagaccccttgcgcttcgacagcctcccgcggttcttggtctcc	1549		
Db	329	AGAAAGACCTCGACAGACAGCCCTCTCGCTTGACGCGCTCCCGCGGTTTGCGCTTC	388		
OY	1550	cgaggtctctctcccgagcgccacacgcttgcaagattgaactcgaacttggcagcgcgcg	1609		
Db	389	CGGCTTCTCTCTCCGGGCGCCACCGCTGAGAGTTGACTGTGAGCTTGGCGACCGCGCG	448		
OY	1610	gtctcaagcttgagggttgagccgaggagggttgaggagacagaggagatgtagactaagc	1669		
Db	449	GCTGCAGTGTGGGGGTGCGCCGGGAGGGGGTGGAGGAGACAGGAGATGTGAGCTCAGC	508		
OY	1670	gcgaggaagcggtctgagccggtgatactatctgcaccaagcagtgctggcgagcaact	1729		
Db	509	GCCGAGGACGCGCTGTGGGGCGCGTGAATCATCTC--GCACCAGAGTGTGCGGCCAGCACT	566		
OY	1730	ccccgggacacagacactgctgcgtgagcagatcccgcgcaagcgttgagagtgcgacctga	1789		
Db	567	CCCCGGGACACGACCTG--CCGCTGACCGAGATCCCGGCG--GGCTGAAGATGCACCTCGGA	624		
OY	1790	ctacgaagcgaggacagtgtagccctccacaagcccaagagcccaagggcgcaactccac	1849		
Db	625	CTACGAGAGCGGGGCAAGTGAACCTCCACAAACCCAGAA--CCAGSAGGCCATCTTAC	662		
OY	1850	ttaactggtcttcttctccgccaaggtlctctctctcttgccgccttgacacaaag	1909		
Db	683	TTTCACTGGCTCTTTCTTCGGGCCAAAGTCTTCCCTGCTCTTGCGCGCTGTGACACAAAG	742		
OY	1910	gtctcgtgccttaagctgcagacagcgaggaaatggggcgcgcggaagggcgagcgaagcgagac	1969		
Db	743	GTTCTTGCTCTTAAAGCTGACACAGGGGGGAAATGGGGGCGCGCAAGAGCGCCAGCCGGGACA	802		
OY	1970	ggcggtcctccgagatccag 1989			
Db	803	GGGGGGCGACCGGGCACACAG 822			

ACCESSION	MRNA sequence.
VERSION	BC386415
KEYWORDS	BC386415.1 GI:13279861
SOURCE	EST.
ORGANISM	human.
	human.
	Homo sapiens

REFERENCE	1 (bases 1 to 952)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: LC6M308 row: b column: 12
High quality sequence start: 7
High quality sequence stop: 586.

FEATURES	Location/Qualifiers
source	1. .952

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TMACE:4583723"
/clone_lib="NH_MCC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the Laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```

Query Match	21.3%;	Score 555.4;	DB 11;	Length 952;
Best Local Similarity	86.8%;	Pred. No. 2.5e-105;		
Matches 723;	Conservative 0;	Mismatches 81;	Indels 29;	Gaps 9

OY	351	cctccatgtcctctctctctctctctgctctcaagcttgaacccggtttcccctccctcg	416
Db	8	CTCCTCTCTCTTTCCTCTCTCTGCTCTCAAGCTTGACACCCTTTCCCTCTCGA --	64
OY	411	ctcaagtgttagtgtagactgcattcaaggaaagaagacctgcgttggtctcgaagagaccg	470
Db	' 65 --	-TTCATGTTAAGTGTGACTGCATTTCAGGGAAAGGAACCTGCGTGGGCTGAGGAGCCG	122
OY	471	gaqfgyaacgggctltygggaaagacacgltgatlgccccacaaccccgctccctctygaagaygtg	530
Db	123	GAGTGAACGGGCTGGGGAGGACACGATGATCCGCCGAACCCCGTCC ----TGAAGTG	177
OY	531	gtccaataagctgctgctgttaacctctgtfyggggcgcgcttgaagagatgcygtgacatt	590
Db	178	GTCCATGAGCGCTGTGCTGTACCCCTATGTGCCCCGGGCGGTGGAGAGTGGGTGACCATTT	237
OY	591	ccctgttgaaacacacctctctgcggcctctgacctcccgcgctctccagatfgyggggccaa	650
Db	238	CCCTGTGGACACACACTTCTTGCCGGGCTTGCTTCCCCTCCGCTCTCCAGATGGGGCCCAA	297
OY	651	tccctctfygcaagaatccctgtcctctcccgctctgcgaagagagtlagacagycagaagacttc	710
Db	298	TCTCTCG--GGCAAGATCCTGCTGTGCTCCGGCTCTGTGCAAGAGGAGGACGAGCAGACTCC	356
OY	711	catgagccctgtgctccctbgggcccgcttgaagataactaatctgcagagagacagcgagaa	770

```
|||||
Db 357 CATGCCCTGTCGCCCTGGCCCGCTGGAGAG-AACCTTACTGCGAGGACGCGGAGAGA 415
Qy 771 agatctactctctcttgagagaagatgcccagttctctctgtgtcttgcagggagaggt 830
Db 416 ACATCTACTTCTTC-TGCAGAGAACGATGCCAGTTCCTGTGTCTTGTGACGGAGGGGT 474
Qy 831 ccacagcaccagcgcacacccgttggtctcctggaagagccattagccctaccggagat 890
Db 475 CCCACGCCACAGGCGCACACCGTGGGGTTCCTGAGACGAGG-CATTACGCCCTACCGGGAT 533
Qy 891 cgtctcaagagtcgaacttggaagctctgagcagcggagagatgagatgtgtaagtgttaa 950
Db 534 CCTCTCAGAGTGCAGCTGAGAGCTGACGACGAGAGATGAGATGAGATGATGATGATGAT 593
Qy 951 ggtctcaagagaccagaagcttcaagtgctgctgactcagatcagcaagaagaagccg 1010
Db 594 GTCAAAAG-CCAACTTCAACTGCTGTGACTCAGATCGAACAAGCAAGAGACCG 648
Qy 1011 tcaaggtgacacagctcctctgagagctgcaagcggagcggagcagcagcagatgct 1070
Db 649 TCAGGGGTGACACAGCTCTTGAAGAGCTGCAAGCGGAGCTGACACAGCAGATGTC- 707
Qy 1071 cctgtcgcgcagagctgagtgtaagctcgcgagatcacaagatttggaaggagagatga 1130
Db 708 ----CTCCTGACGGCTAGAGCTGAGGAGCGCTGACAGCAGAGATGAGGAGGATCA 763
Qy 1131 atatatcaaaagctctctgaggaagatcaccggcttgagccccaagctcaag 1183
Db 764 TTTT-----CAAAGGTTCTGAGAACCAACCCCGGTTGAGCCCGCGTAAAG 811
```

RESULT 3
AW969591/c 621 bp mRNA EST 01-JUN-2000
LOCUS AW969591
DEFINITION EST18168 MAGE resequences, MAGK Homo sapiens cDNA, mRNA sequence.
ACCESSION AW969591
VERSION AW969591.1 GI:8159435
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 621)
Hedge, P., Qi, R., Abenathy, K., Dharap, S., Gaspard, R., Gay, C., Holt
I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeaman, T.J. and
Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Contact: John Quackenbush
COMMENT Unpublished (2000)
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@ligr.org
Plate: 267
Seq primer: Forward.
FEATURES
Location/Qualifiers
source 1..621
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGK"
/note="Vector: pBluescriptSKM"

BASE COUNT 165 a 155 c 144 g 157 t
ORIGIN

Query Match 16.7%; Score 437; DB 10; Length 621;
Best Local Similarity 97.5%; Pred. No. 9.8e-81;
Matches 465; Conservative 0; Mismatches 10; Indels 2; Gaps 2;
Qy 2134 tgaagctctgagctctctctgacactcagagtgacagaccagcgcttcgctgtgc 2193

```
|||||
Db 619 TCAGCTTTCGACATTATATATGCCCTCAGAGTACAGAACTCAGAGGGCTTCGGCTGAGCC 560
Qy 2194 tagggcaaaagccaaactaggaaccgcccgtcttctggggaanaactaaagaagagac 2253
Db 559 TAGGGCAACAGCCAACTAGAGAGCCAG-CGGGCTTTCGGGGAATAAAGAGAGAC 501
Qy 2254 atctaaatgtaatgtttaactgtttcaagataattactctggggaanaactaagggttt 2313
Db 500 ATCTAAATATGTAATGTTTAACTGTTTCAAGATAATTATCTTGGGAAAATACAGGCTTT 441
Qy 2314 gcttgacttgcactaatltgtacagttcaacttcgtacttgcacacacactgaagatgcc 2373
Db 440 GCTGACCTTGCACTAATTTGTACAGTTAAGTTCGTAATTTGACACACACTGAAGATGCC 381
Qy 2374 tccaccttgtaggagcttagggcctttatataagcccttggtgagacccagggcccttc 2433
Db 380 TCCACCTTGTAGGGGCTTGTAGGGCTTTTATCAGCCCTGGGTGAGACCCAGGGCCCTTTC 321
Qy 2434 cttccctccctctcgtgcatcttctcgtacttgtagaagatgttccctaagaagtgtga 2493
Db 320 CTTTCCTTCCCTTCTGTCATTTCTCTGACTTGTAGAAATGTCCTAAGAAAGTGTGA 261
Qy 2494 ctacagaccttgatltccatgtgtccaaatagcgtatgagagctgagagaagagctta 2553
Db 260 CTCACAGACCTCTGATTCATGTCATGTCATTAAGCGCTGATGGAGCTGAGAAAGCTTA 201
Qy 2554 aatcaatggatctgctgtgtgttggaatlttagggccgagatgctcagaggaagt 2610
Db 200 AATCCAAATGGGATC-TGCCTGTGTGTCATTTAGGCCGAGATGCTCAGAGGAGT 145
```

RESULT 4
A1732648/c 666 bp mRNA EST 13-DEC-1999
LOCUS A1732648
DEFINITION nm73f01.x5 NCI_CGAP Co9 Homo sapiens cDNA clone IMAGE:1073881 3'
similar to TR:Q12899 Q12899 ACID FINGER PROTEIN. ;, mRNA sequence.
ACCESSION A1732648
VERSION A1732648.1 GI:5053761
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 666)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaphs-remail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
www-bio.llnl.gov/dbfp/image/image.html

FEATURES
source 1..666
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1073881"

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: National Cancer Institute, Cancer Genome
Anatomy Project (CGAP), Tumor Gene Index
This read has been verified (found to hit its original self in the
correct orientation)
Insert Length: 1309 Std Error: 0.00
Seq primer: -40up from Gibco
High quality sequence stop: 462.

```

/clone_lib-"NCI.CGAP.Co9"
/tissue_type="colon tumor RER+"
/lab_host="RDH10B"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
RER+ colon tumor, and was then primed with a Not I -
RER+(GT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo (Soares4
). "

```

Query Match	16.6%	Score 433.6	DB 10	length 666
Best Local Similarity	88.2%	Pred. No. 5e-80		
Matches 598	Conservative	0	Mismatches 61	Indels 19
				Gaps 11

QY	1612	ttcacggtgggggttggccggggagggggttgagagagacaggagagatgtagtactcagcgc	1671
Db	666	TGCACGCTGGGGGTGGCGGGGAGGGGCTGCAAGGAAGA - ANAGATGGACTCAGCGC	608

[illegible]

QY 1731 ccgggacacgacctgtccgctgagcagatcccgccgacgagcgtgtagagatccgccctgac 1790

	tacagaggcgggcaggttaccctccacaacgcccagagccccaggggcccatcctcaact	1850
OY 1791		

Db	491	TACGAGGGGGGAGGTACCCCTCCACAACGCCAG--CCGAGGAGCCCATCTTCACT	434
Oy	1851	tcaatgagctcttttctccgagcaagtgcttccctgtccttgccgagccgtgacacaaagq	1910

[illegible]

Db 378 TTCCGCTTACGCTGA AAGCTGAAGTGGGGCGCGCGAAGGCGCGGAGACG 320

Db 319 GCGGCTCTCCGGGATCCAGCTCCGGCCCTGGCCAGTGTGGCGGCCGGGGGGCTCCCTTCGTC 260

Accession	Sequence	Position
QY 2031	ccgcgtgagcagagaaacacagcgactgctcgaacagcggtgttttactta	2090
Db 259	CCGCGTgAGCGGAGAGAGA--CAGGGGCACTTAgTCTCAACACAGCGGTTGTTTACTTTA	202

[illegible]

2151 ctgcacctcagtagtgcagaaccacagcgccttcgctgtgtcctagggcaacagccaacc 2210

2211 taggaaccgcgcgcttcgaggaaaactaagaagagacactaanaatgtaatt 2270

Db	81	TAGGAGCCAGCGGGCTTCGGGGGAAAAA----	AAAGAAAAAGACACTTAATAATAAAATGTT	25
Qy	2271	taaacgtttccaagataa	2288	

Db 24 TAACTGTTAAAAAAA 7

RESULT	5
AI801356/c	
LOCUS	
DEFINITION	AI801356 536 bp mRNA Homo sapiens CDNA clone IMAGE:2185352 3

similar to SW:A33_PLENA Q02084 ZINC-BINDING PROTEIN A33. ; mRNA
sequence.
ACCESSION A1801356
VERSION A1801356.1 GI:5366828
KEYWORDS EST.
SOURCE human.

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 536)

AUTHORS NCI-CCGAP <http://www.ncdi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgabbs_r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D.
Michael R

Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/ibrr/image/image.html

Insert Length: 1411 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 424.

```
source 1. 536
        /organism="Homo sapiens"
        /db xref="taxon:9606"
```

```

/clone_1b="NCI CGAP Gas4"
/tissue_type="poorly differentiated adenocarcinoma with

```

```

/lab_host="DH10B"
/note="Organ: stomach; Vector: PCMV-SPORT6; site_1: Sal
site_2: NotI; Cloned unidirectionally. Primer: Oligo d

```

BASE COUNT	Average insert size 1.09 kb. Life technologies catalog			
	11549-011*			
106 a	156 c	165 g	109 t	

Query Match	14.1%	Score 368	DB 10	length 536
-------------	-------	-----------	-------	------------

Best Local Similarity	89.15	Freq. NO. 2.1e-66
Matches	490	Conservative
	0	Mismatches
	45	Indels
		Gaps

Db 535 CCCCCGACCGACCTG-CCGCTGAGCGAGATCCGGCGC-GGCGTGAGAGTGCCTTGA 478

477 CTACGAGCGGGGCGAGTGACCCCTCCACACAGCCACAGA--CCACAGAGCCCATCTTTACC 420
 1750 CTACGAGCGGGGCGAGTGACCCCTCCACACAGCCACAGA--CCACAGAGCCCATCTTTACC 1845

[illegible]

QY 1910 gfocttgcccttagagctgacacgcggggaatggcgcgcgaaaggcgcgcaagcggagac 1969
|| | ||||||| ||||| || ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1970 99cgcctccgcgagatccaaagctccgcgccctggccagctgtcgagcccgaggcctccctgtg 2029

DB 305 GGGGGCTTCGCGGAGTCCACATCCCGCCCTGAGTGTGCGGGGCTCCCTGTG 246

2090	attatcttaggcctcaagctccctctgaagctccctgaagcctccctctgaagcctctgagccttc	2149	
Db	245	cccgcgctgagcgagagAACAA--GGGGACTTGTAGTCTCGAAGACAGCGGTTGTTTTTACTTT	188


```

187 ATTATCTTAGGCCCTCAGCTCCCTGAGCTCCTGAGCCCTCCCTGTGAGCCTTGCCCTTC 128
2150 tctgacctcagatgacgaacacagacgcttcgctgctgctcctgaggaacacac 2209
127 TCTGTACTCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 68
2210 ctgaagaacccgacgcttcgaggaacacacacacacacacacacacacacacacac 2269
67 CTAGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 11
2270 ttaactgt 2279
10 TTAACGTGT 1

RESULT 6
A1222799 525 bp mRNA EST 02-FEB-1999
LOCUS gp39b10.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1925371 3',
DEFINITION mRNA sequence.
ACCESSION A1222799.1 GI:3805002
VERSION A1222799
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 525)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1145 Std Error: 0.00
Seq primer: -40bp from Gibco
High quality sequence stop: 451.
FEATURES
source
Location/Qualifiers
1..525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1925371"
/clone_lib="NCI_CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dt) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 106 a 151 c 157 g 111 t
ORIGIN

```

```

1808 gaccctccacaagccagagcccaagcccaacacacacacacacacacacacacacacacac 1867
465 GA-CCTCCACAGGCCCCAGA--CCAGAGAGCCCATCTTACCTTACGCTCTTT--CT 411
1868 cggccaaggtctccctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1927
410 CCGGCAAGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 354
1928 caaggaggaaatgagggcgagcgaagggcgagcgaagggcgagcgaagggcgagcgaagggcg 1987
353 -AAGGCTGAAGTGGGGCGCGCGCAAGGGCGCGCAAGCGGAGCGGCGGCTCTCGGGATCC 295
1988 agctcgccccggcgaggtgtcgccgcccgggggctcccgctcccgccgctgaggaaggaaga 2047
294 AGCTCGCCCCCTGGCGAGGTGTGGCGCGCGCGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 235
2048 aacacggagactgtgctcgcagacagcgtgttcttactatctatctatctatctatctatct 2107
234 A--CAGGGGACTTGAGTCTCGAAGCGGCTTTTACTTATTATTATTATTATTATTATTATTATTATT 177
2108 gctccctgacgtctcgtgagcctcctcctgctgagcgtctgctcctcctcctcctcctcctcct 2167
176 GCTCCCTGACGTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCTCTGAG 117
2168 gaaccacagagcgtctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 2227
116 GAACCCACAGCGGCTTGGCTGTGCTTGGCGAAGCGGCTTGGCGAAGCGGCTTGGCGAAGCGGCTT 57
2228 ttcggggaataaactaagaagagacatctaaatgtatgtttaaactgttcaagat 2286
56 TCGGGGAAAAA--AAAGAAAAAGACATCTAAATTAATGTTAACTTCAAAAT 1

RESULT 7
A0820993 450 bp mRNA EST 17-MAY-2000
LOCUS RC2-ST0301-240300-016-d01 ST0301 Homo sapiens cDNA, mRNA sequence.
DEFINITION
ACCESSION A0820993
VERSION A0820993.1 GI:7913987
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 450)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?file=st2-RC2-ST0301-240
300-016-d01&3=2000-03-24&4=1)
Seq primer: puc 18 forward
High quality sequence start: 36
High quality sequence stop: 450.
FEATURES
source
Location/Qualifiers
1..450

```

Query Match	13.4%	Score 351;	DB 10;	Length 450;
Best Local Similarity	91.8%;	Pred. No. 7,Le-63;		
Matches 416;	Conservative 0;	Mismatches 30;	Indels 7;	Gaps 4;
Db 683	gcccaaggagtagcagcagcagagactcccatgtgccctgtgcccctgggcccgcgtggag	742		
Db 450	GACAAAGAGAGAGAGAGAGTACAGACAGACATCCCATGCCCTGTGCTGGGCCCCGGTGGAG	391		
Qy 743	ataactactgtcgaaggagcagcgcgagaaagactacttcttcttgcgaaacgatgcga	802		
Db 390	ATAACTTACTGTGGAGAGCAGCGCAGAGAGATCTACTCTTTC-TGCGAAGACGATGCCCA	332		
Qy 803	gtctctctgtgtctctgtcgaaggaggtccacacacacagcagcgaacacgttggttct	862		
Db 331	GTTCTCTGTGTGTCTGTGAGGAGGTTCCACGACACGAGCGCACACCGTGGGTTCT	272		
Qy 863	ggacgagagccattcagccctaccgagatcgtctcaggaagtgcactggaagctctgaac	922		
Db 271	GGACAGAGGCAATTCAGCCCTACCGGATCTCTCAGAGATCTACACGGAAGCTCAGAC	212		
Qy 923	ggaagagatgtgagatgtgtgagatgaaagtgcgaagaagcagaagctcgaagtgcag	982		
Db 211	GGAGAGAGATGTGATTTGTGATGTAAAGTGTCAAGAGAACCAAGACTTCAAGTGGCG	152		
Qy 983	tgactcagatcgacaagaagccgctcgaaggtgcacacacagactctcttgaaaggtcga	1042		
Db 151	TGACTCAGATTCG--CAAGCAAGAACATCATCGCTGTAAACA--GCTTTGAGAGGCTGC-	96		
Qy 1043	agcggagagctgagcagcagcagatctctcgtctgcgagagctgagtgagcctcg	1102		
Db 97	ACCAAGAGCTGGTGCAGACGCCATGCTCTCTGCGGAGGAGCTGAGTGGTACCTGG	38		
Qy 1103	agtcacagatttggaaggagagatgaatga 1135			
Db 37	AGTCACAGATTGGCAGAGTGAAGGAGCTGTATA 5			
RESULT 8				
AK007445				
LOCUS	AK007445	1779 bp	mRNA	HTC
DEFINITION	Mus musculus 10 day old male pancreas cDNA, RIKEN full-length			
ACCESSION	AK007445			
VERSION	AK007445.1	GI:12841001		
KEYWORDS	CAP trapper.			
SOURCE	Mus musculus (strain:C57BL/6J) 10 day old male pancreas cDNA to			
KEYWORDS	mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library			
clone:1810012B10.				
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
TITLE	1 (bases 1 to 1779)			
JOURNAL	Carninci, P. and Hayashizaki, Y.			
MEDLINE	High-efficiency full-length cDNA cloning			
PUBMED	Methods in enzymology. 303, 19-44 (1999)			
REFERENCE	99279253			
	10349636			
	2 (bases 1 to 1779)			

AUTHORS	Carinciuc,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL MEDLINE PUBMED	Genome research. 10 (10), 1617-1630 (2000)
REFERENCE	20499374
AUTHORS	11042159 3 (bases 1 to 1779) Shibata,K., Itoh,M., Aizawa,K., Nagoka,S., Sasaki,N., Carinciuc,P., Kono,H., Akiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matsubiki,M., Yoneida,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kiru,A. and Hayashizaki,Y.
TITLE	Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL MEDLINE PUBMED	Genome research. 10 (11), 1757-1771 (2000)
REFERENCE	20530913
AUTHORS	11076861 4 (bases 1 to 1779) The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL MEDLINE PUBMED	Nature 409, 685-690 (2001)
REFERENCE	5 (bases 1 to 1779) Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Akeawa,T., Carinciuc,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirota,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Itawa,M., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S., Kuhihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Oikado,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakii,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Teijima,Y., Toyota,T., Yamamura,T., Yasunishi,Y., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGGAAGAATCCACAGAGCTCCTTTTTTTTATTGTA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAAGAGAGAGCGGCCGCATTTAATTCGTCAAGTTAATTAAATTAATCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR. * Location/Qualifiers 1..1779 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /db_xref="MGD:MGI:1901147" /db_xref="MGI:MGI:1916347" /cloned="1810012B10" /sex="male" /tissue_type="pancreas" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="10 day old"
BASE COUNT	481 a 447 c 492 g 359 t

Db 129 CCCTGTGCCCGCTGAGCGCAGAGAA--CAAGGACCTTGAGTCTCGAACAACGGCTGT 186
OY 2083 ttacttattatcttagagccctcagctcctcagctcctcgaagctcctcgttgaagctct 2142
Db 187 TTACTTTATTTATCTTAGGCGCTCAGCTCCCTGACGTCCTGAGCTCCTCCTGTGACGCTCT 246
OY 2143 ggcctctctgcacacgaagtgacgaacacacagacgcttcgctcgtcgtcgaagac 2202
Db 247 GGCCTTCTGACACTCAGATGAGTGAACACACACAGCGCTTCGGCTGTGCTAGGCGAAC 306
OY 2203 agccaactaggaaccccgccgcttcggtgggaaataacgaaggaagacataaataat 2262
Db 307 AGCAACCTTAGGAACCCGCGCTTCGGGGAACACAAAGAAAGACGTCTAATC 366
OY 2263 gtaatgttaaacactgttcaagaataatccttggaataatacagggcttcgtcgaact 2322
Db 367 AAACGTG---CAACTGTTCAGATATATCTTGGAACAT-AGGGTTTGTGCTGACTT 422
OY 2323 gcaactaa 2329
Db 423 GCACTAA 429
RESULT 10
AM610230/c 517 bp mRNA EST 23-MAR-2000
LOCUS RC2-ST0301-300100-013-h02 ST0301 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM610230
ACCESSION AM610230.1 GI:7314971
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 517)
HCGP <http://www.ludwig.org.br/ORESTES>.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC2&t2=RC2-ST0301-300100-013-h02&t3=2000-01-30&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 45
High quality sequence stop: 517.
FEATURES
Location/Qualifiers
1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0301"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site:1: Sma1;
Site:2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent Application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 92 a 172 c 128 g 125 t
ORIGIN
Query Match 12.3%; Score 320.6; DB 10; Length 517;
Best Local Similarity 88.2%; Pred. No. 1.5e-56;
Matches 442; Conservative 0; Mismatches 44; Indels 15; Gaps 8;

OY 683 gccaaagagtagcagcagagagactcccaatggccctcgtgcgccctggccgcctggag 742
Db 517 GCCAAGAGAGAGACAGACAGACAGACTCCATGGCCCTGTGCCCCCTGGCCCTGGAG 458
OY 743 ataacttactgcagagagcagcagcgaagaatcactctctctcgcagaagatgcga 802
Db 457 A-AACTTACTGCGAGGACAGCGCGAGAGATCTACTTCTTC-TGCCAGAACGATGCCGA 400
OY 803 gtccctctgtgtctcgcagggatcccaacgacccaagcgacacccgttggttcc 862
Db 399 GTTCTCTGTGTCTTGTGCAAGGAGGCTCCACTGACACAGCGACACCGGGGTTCT 340
OY 863 ggaagagccatcagaccctcagcggatcgtcctcaggaagtcgaatcctcgaac 922
Db 339 GGAGAGAGCCATTCAGCCCTACCGGATCTGCTCAGGATGCGATGGAACCTGAGAC 280
OY 923 ggaagagatgagatgtgtagatgtaagtgtcaagaagacgaagctcgaagtcg 982
Db 279 GGAGAGAGATGAGATGTG-AGGATGTAAAGTCTCAAGAGACCAAGAACTTCAAGTCTGC 221
OY 983 tgactcagatcgacaacagcagaagccgtcaagggtgcacacagctccttaagagctgca 1042
Db 220 TGACTCAGATCGTATCCTTAGAAGCATC---AGGTGAGACAGCT-TTGTAGAGAGCTGC- 165
OY 1043 agcggagcctgcagcagcagatgtctcctcgtcgtcgcgaagatgagtgacgtcgg 1102
Db 166 AGCAGAGCTGAGACACACGATGCTCCTGCTGGCCAGGCTGAGGAGACTGGAGACG 107
OY 1103 agtcacagattggaagagagagatgatatatcacaaagctctcgaagaatcaccc 1162
Db 106 AGAC-----TTTGGAAAGAGAGGATGAATATATCACAAAGCTCTGAGGAAGTCAACC 52
OY 1163 ggccttgagccccaagcgcgaag 1183
Db 51 GGCTT-GAGCCCAAGCTCAGG 32
RESULT 11
BF431333/c 449 bp mRNA EST 29-NOV-2000
LOCUS naa43f01.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3259152 3'
DEFINITION similar to TR:095604 095604 ZINC FINGER PROTEIN.; mRNA sequence.
ACCESSION BF431333
VERSION BF431333.1 GI:11443447
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 449)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
infoimage.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 288.
FEATURES
Location/Qualifiers
1..449
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3259152"
/clone_lib="NCI-CGAP_Kid11"

SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 418)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT	Tumor Gene Index Unpublished (1997) Contact: Robert Strusberg, Ph.D. Email: cgaps-remail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center clone distribution: NCI-CGAP clone distribution information can be found through the 1.M.A.G.E. Consortium/LNL at: www-bio.lnlnl.gov/bdbrp/image/image.html Insert Length: 1123 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 404.
FEATURES	location/qualifiers 1..418 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1873192" /clone_1lb="NCI CGAP Co8" /tissue_type="adenocarcinoma" /lab_host="DH10B" /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - o119o(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonalduo."
BASE COUNT	92 a 118 c 117 g 91 t
ORIGIN	
Query Match	11.8%; Score 308.2; DB:10; Length 418;
Best Local Similarity	89.0%; Pred. No.5.6e-54;
Matches 380; Conservative	0; Mismatches 38; Indels 9; Gaps 4;
QY 1860	ctttctccgccaagagcttcctcgtcttccttgcgcgccttgagacacaagaagtctgcgct 1919
Db 418	ctcttttccggcgaagaagcttccctttttgttgcgcttg--AAAAAAGGTTCTGCT 362
QY 1920	tagctgtgacacgggggaaatgtgggcgcgcgaaggcgcgaagcgagaaacggcgcttc 1979
Db 361	TACGCTGA-AAAGCGTGAAGTGGGGCGCCGAAAGGCGCGAAGCGAGACGGCGCTCTC 303
QY 1980	cggagtcacagctcgcgccttgacagctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2039
Db 302	CGGATCCAGCTCCGCCCTCGCCAGTGTGGCGCCGGGGGCTCCCTGTGCCGCGTAG 243
QY 2040	gcgagagaacaacgggagcttgagcttcgaaacagcgctgttttaacttaattactta 2099
Db 242	GCGAGAGAA--CAGGGGCACTTGAGCTCGAACAAGCGGCTTTTACTTTATTTATCTTA 185
QY 2100	ggcctcagctcccttgaagcttgaagcttcctcctgtgaaagctctgagcctctctgaactc 2159
Db 184	GGCCCTCAGCTCCCTGACGCTCCTAGCCTCCCTGTGAAGCTGTGGCTTCTGTGACCTC 125
QY 2160	agaatgcagaacacacagcgcgctctgcgtctgcctcaaggcaacaacacacctaaggaacc 2219
Db 124	AGAGTGCAGAACCAACAGACGGCTTCGGTGTGGCTTAGGGCAACAGCCCACTAGAGGCA 65
QY 2220	gcgcgccttctcggggaaactaaagaagagacactaaatgtaatgtttaactgtt 2279
Db 64	GCGGGCTTTCCGGGAAAAA--AAAGAAAGACATCTAAATATAATGTTAAACTGT 8

Oy	2280	tcaagatc	2286
Db	7	TCAAAAT	1
<hr/>			
RESULT	14		
LOCUS	A1791944		
DEFINITION	nm73f01.y5 NCI-CGAP Co9 Homo sapiens cDNA clone IMAGE:1073881 5' similar to TR:Q62157 Q62157 ZINC FINGER PROTEIN ; mRNA sequence.	EST	13-DEC-1999
ACCESSION	A1791944		
VERSION	A1791944.1	GI:5339660	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 491)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Other_ESTS: nm73f01.x5 Contact: Robert Strausberg, Ph.D. Email: cgapbs@email.nih.gov Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.		
JOURNAL COMMENT	CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA sequencing by: Washington University Genome Sequencing Center clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.lnl.gov/bdrp/image/image.html		
<hr/>			
This read is a RESEQUENCE of a previously sequenced human clone Original clone citation: see original entry for original citation Information This 5' resequenced clone has no previous 5' data to verify this new read against Insert length: 1309 Std Error: 0.00 Seq primer: -40RP from Gibco High quality sequence stop: 462. Location/Qualifiers 1. .491 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1073881" /clone_lib="NCI-CGAP.Co9" /tissue_type="colon tumor RER+ " /lab_host="DH10B" /note="Organ: colon; Vector: pTZ19D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from RER-colon tumor, and was then primed with a Not I - oligo(OT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTZ13 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo (Soares			
FEATURES	source		
<hr/>			
BASE COUNT	135 a	115 c	149 g
ORIGIN			92 t
<hr/>			
Query Match	11.7%; Score 306;	DB 10:	Length 491;
Best Local Similarity	88.8%; Pred No. 1.7e-53;		
Matches 444:	Conservative 0;	Mismatches 20;	Indels 36;
			Gaps 9;
OY	1030	ttgaaggctgacaagcggaagtgcgaagcagcagcatgtctctgctgtagcgaagactga	1089
Db	28	TTGAGAGGCTGC-AGCAGAGACTGGAGCAGCAGCATGTCTCTGCTGCCAGGCTGAGG	86
OY	1090	gttgtaagcttcgaagctcacagaatttggaaaggaaggaatgatataaacaaaagctctct	1149

Db 87 GAGCT-----GGACGACAGATTTGGAAGGAGGAGATATATACAAAGGTCTCT 140
QY 1150 gaggaagtcacccggttggagcccaagtcgaagcctcgagaggaagtgctcagcagc 1209
Db 141 GAGGAGTCACCCGGCTTGAG-CCGAGGTCAAGAGCT-GGAGGAGGAGGTGTACAGCAGC 198
QY 1210 cagcaagtcagcttcacaagaatgcaagtcagaagcccaagcaggtgtgagatgaagact 1269
Db 139 CAGCAAGTCAGCTTCTACAGATGTACAGGTCAA-CCAGAGAGAGGTGTAGTGAAGACT 257
QY 1270 ttgtgagtcctgagcgaatctccctgacctgtlcaagaagatccgtgattccacag 1329
Db 258 TTTGTGAGTCTGAGGCGCTTTCT-CTGTACCTTGTCAAGAGATCCGATTTCCACAG 316
QY 1330 gaaatctacccctcccaagaatgatgaatgttctcaaaaacttgccgcatcacc 1389
Db 317 GAAATATCTACCCCTCCGAGATGATGAGATTTCT-AGAAAACTTGCGCATCATC 375
QY 1390 tggaaatcagcaggggtcatcactctgacccctcagcccgccagccgagacactgtt 1449
Db 376 TGGAAATGAT-----TCAGACCCCGCAGCCGGAG-CTGTGTT 411
QY 1450 ctctcggaagacaggaagtcagtgagtcaccccgagcagaagaagcctgcacagacagc 1509
Db 412 CTCTCGGAGACAGGAGTACGTGACACCCGCGCAGAGAAAGAACTGCCAGACAGC 471
QY 1510 cccctgagcttcgaagcct 1529
Db 472 CCCCTGCGCTTGACGGCT 491

RESULT 15

LOCUS AA527898 450 bp mRNA EST 20-AUG-1997
DEFINITION ne8d09.s1 NCI_CGAP_C03 Homo sapiens cDNA clone IMAGE:901553 3',
mRNA sequence.
ACCESSION AA527898
VERSION AA527898.1 GI:2269967
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 450)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga9bs-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNL at:
www.bio.lnl.gov/btrp/image/image.html
Seq primer: -21m13 forward (Amersham)
High quality sequence stop: 173.
Location/Qualifiers

FEATURES

Source
1. 450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:901553"
/clone_lib="NCI_CGAP_C03"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: p7T73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed

with a Not I - o1lgo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p7T73 vector. Library went through one round of
normalization.

BASE COUNT 108 a 115 c 128 g 99 t

ORIGIN

Query Match 11.7%; Score 305; DB 10; Length 450;
Best Local Similarity 89.0%; Pred. No. 2,6e-53;
Matches 398; Conservative 0; Mismatches 40; Indels 9; Gaps 6;

QY 1903 acaaaggtcccgccctgagctgacacgaggaatggggcggcgaagggcgaggaag 1962
Db 10 AAAAAAGTCTTCGCTTACGCTGA-AGGCTGAAAGTGGGGCCGCGAAGGGC-GGAAAC 67
QY 1963 cggagacggcggcctctccgggaltccagctccgcccctggccagtgctcgcccgaggct 2022
Db 68 GGAGACGGCGGCTCCGGGATCCAGCTCCGCCCTGGCCAGTGTGCGCGGGGGGCT 127
QY 2023 cccctgtcccgctgagcggagagaaacacggggaacttgatctcgaaacgggtgtt 2082
Db 128 CCCTGTGCCCGGTGAGCGAGAGAA--CAGGGGACTTGAAGTCTGAACAGCGGTGTTT 185
QY 2083 ttacttatctatcttagccctcagcctccctgagctcctgagcctccctgtgagcctct 2142
Db 186 TTACTTTATTTATCTTAGGCCCTCAGCTCCTGACCTCTGAGCTCTGAGCCTCT 245
QY 2143 ggcctctctgacctaagagtgacagaaacacagacgcttggctgtgctcaggagcaac 2202
Db 246 GGCCTTCTCTGCACCTCAGAGTGCAGAACACACAGACGGCTTGCGGCTGTGCTAGGG- AA 303
QY 2203 agccaacctaggaaccccgcccttcgggggaaactaaagaagagacatctaaat 2262
Db 304 CAGCAACCTTAGAGGCCAGCGGG--TTTCGGGGGAAAAAAGAAAGAACATCTAAAT 361
QY 2263 gtaatgtttaaactgttttaagaataatcttggggaataatcagaggttttctgagct 2322
Db 362 AAATGCTTAACGTGTTCAA-AAATATATCTTGGGGAACCAAGGGGTTTGCTGAGCTT 420
QY 2323 gcactaattgtacagttaacttcgta 2349
Db 421 GCACATAATTTTACGAGTAACCTTCTTA 447

Search completed: March 30, 2002, 09:32:30
Job time: 14007 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 13:07:47 ; Search time 519.52 Seconds

(without alignments)
4307.089 Million cell updates/sec

Title: US-09-867-034-5

Perfect score: 2610
Sequence: 1 gatgtgcgcagcctccagag.....ccgagatgctcgcaggaagt 2610

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_1101:*

1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	208	8.0	3227	22	AA158226 Human polynucleot
2	192.2	7.4	2351	22	AA160012 Human polynucleot
3	76.4	2.9	520	22	AA123948 Probe #13881 for g
4	76.4	2.9	520	22	AA149253 Probe #17939 used t
5	76.4	2.9	520	22	AA109542 Probe #9533 used t
6	76.4	2.9	3518	19	AAV57905 Bovine butyrophilin
7	69.2	2.7	2981	21	AACT7707 Human cancer assoc
8	67.6	2.6	43226	20	AAK60263 Nucleic acid seque
9	63	2.4	712	21	AAAT8067 cDNA encoding huma
10	63	2.4	712	22	AAAT8805 Colon tumour relat
11	63	2.4	43226	20	AAK60263 Nucleic acid seque

12	60.6	2.3	376	22	AAH72187
13	60.6	2.3	509	22	AAH71127
14	60.6	2.3	750	22	AAH72832
15	60.6	2.3	14180	19	AAV05159
16	59.6	2.3	571	20	AAV87250
17	59	2.3	3479	22	AA521305
18	58.8	2.3	3161	16	AAO80233
19	58.8	2.3	3344	16	AAO80228
20	58.8	2.3	5191	21	AAZ88617
21	58.2	2.2	2280	17	AAI6127
22	57.6	2.2	400	22	AAE64260
23	57.6	2.2	411	22	AAE65013
24	57.4	2.2	243	21	AAH30668
25	57.2	2.2	395	22	AAE64375
26	57	2.2	428	22	AAE64255
27	57	2.2	3515	22	AAE75343
28	57	2.2	5403	20	AAV84140
29	56.8	2.2	387	22	AAE67194
30	56.8	2.2	2904	19	AAV57904
31	56.4	2.2	412	22	AAE66777
32	56.2	2.2	471	22	AAE69324
33	56.2	2.2	49999	20	AAZ23889
34	56	2.1	558	22	AAI16803
35	56	2.1	558	22	AAI39486
36	56	2.1	558	22	AAI40381
37	56	2.1	4122	22	AAAD10239
38	56	2.1	7715	19	AAVA3042
39	55.8	2.1	383	22	AAE64661
40	55.8	2.1	1462	20	AAZ10861
41	55.8	2.1	2095	22	AAAD07708
42	55.6	2.1	386	22	AAE64628
43	55.6	2.1	402	22	AAE64814
44	55.6	2.1	700	11	AAO05263
45	55.6	2.1	700	15	AAQ57505

ALIGNMENTS

RESULT 1	
AA158226	AA158226 standard; cDNA: 3227 BP.
XX	XX
AC	AA158226:
XX	XX
DT	22-OCT-2001 (first entry)
XX	XX
DE	Human polynucleotide SEQ ID NO 429.
XX	XX
KW	Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW	leukemia; ss.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO20015312-A1.
XX	XX
PD	26-JUL-2001.
XX	XX
PE	26-DEC-2000; 2000WO-US34263.
XX	XX
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662192.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	XX

Human cervical can
Human cervical can
Human cervical can
Mouse butyrophilin
EST clone B0260.
Human cDNA sequenc
Rat NDF clone 42B
RPP1-Wsa genomic n
Mouse neuropeptide
Novel human polynu
Novel human polynu
Human colon cancer
Novel human polynu
Novel human polynu
Human rGF-beta rec
Mouse prothrombina
Novel human polynu
Bovine butyrophilin
Novel human polynu
Human cervical can
Human LOBO homolog
Probe #6736 for ge
Probe #8172 used t
Probe #9067 used t
Mouse long whey ac
Mus musculus Cgamma
Novel human polynu
Hyaluronate syntha
Human secreted pro
Novel human polynu
Novel human polynu
Sequence of neuron
Rat GAP-43 promote

PR	09-JUL-2000;	2000US-0598042.	
PR	19-JUL-2000;	2000US-0628312.	
PR	03-AUG-2000;	2000US-0634550.	
PR	14-SEP-2000;	2000US-0662191.	
PR	19-OCT-2000;	2000US-0693036.	
PR	29-NOV-2000;	2000US-0727344.	
XX			
PA	(HYSE-) HYSEQ INC.		
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang Z, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX			
DR	WPI: 2001-443253/47.		
DR	P-PSDB; AAM40856.		
XX			
PS	Claim 1; SEQ ID NO 4001; 10078bp; English.		
XX			
CC	The invention relates to human nucleic acids (AA157799-AA161369) and		
CC	the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: Immune system suppression,		
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, leukaemias and		
CC	C.N.S disorders.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification.		
XX			
SO	Sequence 2351 BP: 577 A; 575 C; 680 G; 518 T; 1 other;		
Query Match			
Best Local Similarity 56.1%; Score 192.2; DB 22; Length 2351;			
Matches 667; Conservative 0; Mismatches 493; Indels 29; Gaps 15			
OY	715	gcccctgycgccctgycgccgcgtgagagataactactctcgagagagacgycgagagaat	774
DB	271	gctggctccacactgaggtttggagagagagatgtctccaaagacacgagagagaat	330
OY	775	ctactctctctgagagacgaatgcgcgaattcctcctgctgtgtcttcgagggaggtccca	834
DB	331	ctactctctc-tgttagagatgattgagatgcaattgtgcgtgtgtgycggagagctggg	389
OY	835	cgcacagcgagacacgcgttgggttctctctgacgaaggc-caticagccctacccggagatc	893
DB	390	agcagctacccacacccatgccttccctgagatgacgcygctccctataggagacaa	449
OY	894	ctcagagtgatgactggaagctctgagcagcgagagagatgaatgttaagtgtaaag	953
DB	450	atccataagtgtcttaaaatgcttaataaaagagagagagagatt-caagaatccagat	508
OY	954	tcaagaagaaccgaagctcaagtgcgctgtgactcgatccgaacgaagcagaagccgtca	1013
DB	509	aaggaataaataaagatgcaagctcctcctgctcaggtcgtccaccaaagacaaagat	568
OY	1014	gggtgacacacagctctcttgagagagcgtgcaagcgggagctgcagcagcagatgttcct	1073
DB	569	gattcttga-----gttcgacacaccttgagggaattctcagaggaacacagcagcatcct	622
OY	1074	gctgscgagagctgagtgctacgctcgtgacgcacagatcttgagaggaagagatgaata	1133
DB	623	cttagcacac-----atggagagccaggaatggtggacacattctgaagcaacgagatgaatt	676

QY	1134	tatccacaaggtctcttgaggagatccacggcttggagccccaactcaagagactcgag	1193
Db	677	tgaattgcttggttgcctggaggagatctgcggttttagtctctta ttgaagact -	734
QY	1194	gagaaagtctcaagcaacgcagcaagtgaagcttcaaaagtgccagagttcaagccag	1253
Db	735	gggaagaatgaagaagccacagcaagggagatccctgaagacacaaagaactcta -ataag	793
QY	1254	gtctgagatgaagaacttttgtgattctcgaaggccattctctccgactgttccaagaa	1313
Db	794	atgtgaacaacagaaagtgcggaaaccggtgctgtgtgc -gccagagcttggccaagsga	852
QY	1314	tccgtgatcttccaagagaataatactcacctccacagagatga tgaagaatgtctcaagaa	1373
Db	853	tctcgagactttccccaagcagccctcccgctgcagagagagatgaagaatgtt -cttgaa	911
QY	1374	aacttggcgcatcatcttgaaatgaattgaattcaagggtcatcatcttggaaacctcaagcc	1433
Db	912	aaactatgctttgaatttgagactatgaatagccactcaacattctcagaaacctcaacttc	971
QY	1434	agccggagagactgtgtctctctcgaaagacaggaagtgcaatgagatcaaccgycagaa	1493
Db	972	caccacca -agctccctctgtctcgagagaccacacagcagatcaagttcttccataaatgca	1030
QY	1494	gagcgtctccagaacagccctctgccttcgacagccctcccgcgatctctgagcttccgg	1553
Db	1031	gaactcaccaagacaaccccccaaggtttgaaccgggcacactgttcttgcaccaatgg	1090
QY	1554	ctctccctccggcgcccaaccgctgcaagttgaactgaagctggcgagcggcgctg	1613
Db	1091	catccaaagggggagacacacacagctgggtggatgaataagacctggccaatggggcagctg	1150
QY	1614	caaggctggggggcgccggggaggggtgaaggagacaggaagagatggaactcagcgcg	1673
Db	1151	cacgctggcggtggtgagacgaagaatgtgcacgga -agggggaaatcttcgctgcggccag	1209
QY	1674	aggacgagctctgggcgtgatacatctctgcacaaagcagtgtcgggcagcacctccc	1733
Db	1210	aggaggggggtggggctgtgaagctggttgggcttc -gtctcgactctgctcttc	1268
QY	1734	gggacacgaactgtccgctgaagagatcccgcgcaagcggtgaagctgcgacctgaactac	1793
Db	1269	ccccacacagctgacacctgaaggagacgcccggcgag --gtgaagggttctctgaactat	1326
QY	1794	gagcgggcgacaggtacacctccaacagcccagagggccacatccctcaactca	1853
Db	1327	gagggcgggctgggtgaacttcaacagcgtctacccggagag --cccatcaacacttca	1384
QY	1854	ctgagctcttctccgycgaaggtcttccctgcttcttggccgcgactgac	1902
Db	1385	ctgctctcttcaactag --aaggatcatcctcttcttgggtctcggggc	1431
RESULT 3			
ID	AAI23948/c		
XX	AAI23948 standard; DNA: 520 BP.		
AC	AAI23948;		
XX	12-OCT-2001 (first entry)		
DT	Probe #13881 for gene expression analysis in human cervical cell sample		
XX			
DE	Probe: human: microarray; gene expression; cervical epithelial cell;		
XX	cervical cancer; SS.		
KW			
XX	Homo sapiens.		
OS			
XX	WO200157278-A2.		
PN	09-AUG-2001.		
XX			
PD	30-JAN-2001; 2001WO-US00670.		
XX			

```

XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 13881; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 520 BP; 111 A; 158 C; 137 G; 114 T; 0 other:
XX
Query Match 2.9%; Score 76.4; DB 22; Length 520;
Best Local Similarity 57.4%; Pred. No. 8.3e-10;
Matches 218; Conservative 0; Mismatches 151; Indels 11; Gaps 4;
XX
OY 1441 gacctggtctctcgaaagacagaagtcagtgaggtacacccgcgaagaagagctcg 1500
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 413 GAGCTCTCTCTCTGTCAGAGACCGGAGAAAGTGTGAGGCGGGCCCTACAGCGAGAGTG 354
XX
OY 1501 ccagacagcccccttcgcgttcgaagccctcccgcggtcttggtctccgggtctcc 1560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 353 CCTGCAACCCCGAGAGATTCGACAGTCAAGCTTGTGTCTTGGAATGGAGAGCTTCGCC 294
XX
OY 1561 tccggcgccacagcgtcgcaaggttgacctgcagctcgcgacgcgcgcgctgcaggtg 1620
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 293 TCAGCGAAACATTAATTACTGGGAGG-----TGGAGGTGGAAACGTGATGTGTGACCTGTG 240
XX
OY 1621 gggcgcgccggggaagggttgagagggagacaaaggaaagatggagatccagcgccgaagagcg 1680
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 239 GGGGCTCTCAGACACAGTGTGAGAGGAAAGG--GAGGCTCGTGTGATTCTCTCAAAATGG 181
XX
OY 1681 cgtcggggcgtgatcatctctgcacccaagcagtgctgtgcgcagcacctccccgggacac 1740
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 180 CTTCTGGAACCGTCGAGAGATGTTTGA--AACCAATACGGGGCCCTTCTCTCCCTG--AGA 125
XX
OY 1741 gacctgtccgctgagcagatcccgcgcaagcgctgagagtgccctgactacgaagcgcg 1800
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124 GGATTCCTCCCTTTGAAGGAGTCCCTTTCGCGGTTGGGCGTCTTCTCGACATGTGAAGCTG 65
XX
OY 1801 ggcagctgacctccacaac 1820
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 64 GAGATGTCCTCTTACAAAC 45
XX
RESULT 4
AA149253/c
XX AA149253 standard; DNA; 520 BP.
XX
XX AA149253:
XX

```

DT	17-OCT-2001	(first entry)
XX		
DE	Probe #17939 used to measure gene expression in human placenta sample.	
XX		
XX	Probe; microarray; human; placenta; antenatal diagnosis;	
KW	genetic disorder; ss.	
OS	Homo sapiens.	
PN	WO200157272-A2.	
XX		
PD	09-AUG-2001.	
XX		
PF	30-JAN-2001; 2001WO-US00663.	
XX		
PR	04-FEB-2000; 2000US-0180312.	
PR	26-MAY-2000; 2000US-0207456.	
PR	30-JUN-2000; 2000US-0608408.	
PR	03-AUG-2000; 2000US-0632366.	
PR	21-SEP-2000; 2000US-0234687.	
PR	27-SEP-2000; 2000US-0236359.	
PR	04-OCT-2000; 2000GB-0024263.	
XX		
PA	(MOE-) MOLECULAR DYNAMICS INC.	
XX		
PI	Penn SG, Hanzel DK, Chen W, Rank DR;	
XX		
DR	WPI; 2001-488897/53.	
XX		
PT	Human genome-derived single exon nucleic acid probes useful for	
FT	analyzing gene expression in human placenta -	
XX		
PS	Claim 25; SEQ ID NO 17939; 654pp; English.	
CC		
XX	The present invention relates to single exon nucleic acid probes (SENP).	
CC	The present sequence is one such probe. The probes are useful for	
CC	producing a microarray for predicting, measuring and displaying gene	
CC	expression in samples derived from human placenta. The probes are useful	
CC	for antenatal diagnosis of human genetic disorders.	
XX		
SQ	Sequence 520 BP; 111 A; 158 C; 137 G; 114 T; 0 other;	
	Query Match	2.9%; Score 76.4; DB 22; Length 520;
	Best Local Similarity	57.4%; Pred. No. 8.3e-10;
	Matches 218; Conservative	0; Mismatches 151; Indels 11; Gaps 4;
OY	1441 gaccggttcctcgaagacagagatcagtgaqtacaccggcagaagaagacctg	1500
Db	413 GAGCCTTCCTGTCAGAGACCGGAGAACTGTGAGCGGGGCCCTTAAGGCAGAGACTG	354
OY	1501 ccagacagccccctgcgtcgtcgaagccctcccgagcgttcttggtctccgggtctcc	1560
Db	353 CCTGCACAACCAGAGAGATGTGACAGTCAAGCCTTGCTCTGGGATGGAGAGCGTTCCGC	294
OY	1561 tcggaggcccaacgcgtgcgaattgacctgcagctggcgagcgcgagctgtgacagtg	1620
Db	293 TCAGGGAACAATTATCGGGAGG-----TGAAGGTGAAAACTGATGCTGTGACTGTG	240
OY	1621 ggaggctggccgggaggggtagagagacagagagatactgagactcagcgccgagagacg	1680
Db	229 GGGGTCCTCAGACACACAGTGTGAGAGAAAGG-GAGTGCTCGCTGATTTCTCAGAATATG	181
OY	1661 cgtctggcgctgatcatctctcgaaccaagatgctgctggccaacaactcccgggcaac	1740
Db	180 CTTCCTGGACCTCTGAGAGATTGGA--AACCAATAACCGCGCCCTCTCCCTCG--AGA	125
OY	1741 gaccgtgcccttgagcagagatcccccggagcgctgagagatcgcccttgactacagagcg	1800
Db	124 GGATTCTCCCTTTGAAGGAGTCCCTTTCGCGGCTGGCGCTTCTCTGACTATGAAGCTG	65
OY	1801 ggcagttgacctccacaac	1820

Db	64	GAGATGTCCTCTTCAAC	45
RESULT	5		
ID	AAI09542/c		
XX	AAI09542	standard; DNA: 520 BP.	
AC	AAI09542;		
XX			
DT	09-OCT-2001	(first entry)	
XX			
DE	Probe #9533 used to measure gene expression in human breast sample.		
KW	Probe; human; breast disease; breast cancer; development disorder; ss;		
XX	Inflammatory disease; proliferative breast disease; non-carcinoma tumour.		
OS	Homo sapiens.		
XX			
PD	WO200157270-A2.		
XX			
PF	09-AUG-2001.		
XX			
PR	29-JAN-2001; 2001WO-US00661.		
XX			
PR	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024265.		
XX			
PA	(MOLE-) MOLECULAR DYNAMICS INC.		
XX			
PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX			
DR	WPI; 2001-476286/51.		
XX			
PT	Novel single exon nucleic acid probe used to measuring gene expression		
PT	in a human breast -		
XX			
PS	Claim 25; SEQ ID NO 9533; 322pp; English.		
XX			
CC	The present invention relates to novel single exon nucleic acid probes.		
CC	The present sequence is one such probe. The probes are useful for		
CC	measuring human gene expression in a human breast sample, where the probe		
CC	hybridises at high stringency to a nucleic acid expressed in the human		
CC	breast. The probes are useful for predicting, diagnosing, grading,		
CC	staging, monitoring and prognosing diseases of the human breast,		
CC	particularly those diseases with polygenic aetiology. The diseases		
CC	include: breast cancer, disorders of development, inflammatory diseases		
CC	of the breast, fibrocytic changes, proliferative breast disease and		
CC	non-carcinoma tumours.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences .		
XX			
SQ	Sequence 520 BP; 111 A; 158 C; 137 G; 114 T; 0 other;		
XX			
Query Match	2.9%;	Score 76.4;	DB 22; Length 520;
Best Local Similarity	57.4%;	Pred. No. 8.3e-10;	
Matches 218; Conservative	0;	Mismatches 151;	Indels 11; Gaps 4;
Db	413	GAGCTCTTCTCTCAGAGAGACCGGAGAACTGTGAGGCGGGCCCTTACAGGAGAGATG	354
XX			
YY	1441	gaccggtctctctggaagacagagaaatcagtgaagctacaccgcgcagagaagagctcg	1500
XX			
YY	1501	ccagacagccctctgcgtctgcagcggcctcccgagcgtctctggtctcccggtctcc	1560
XX			
Db	353	CCTGACAAACCCAGAGATTCAGACAGTCAGCCTTGTCCTGTGGATGGAGAGCTTCGCC	294
XX			
YY	1561	tcgggagccacacgcgtgtgcaggttgaacctgcagcttggcgacagcgcggtctgcaggtg	1620

Accession	Gene	Species	Length (bp)	GC Content (%)	GC Skew	GC Bias	GC Bias2	GC Bias4	GC Bias8	GC Bias16	GC Bias32	GC Bias64	GC Bias128	GC Bias256	GC Bias512	GC Bias1024	GC Bias2048	GC Bias4096	GC Bias8192	GC Bias16384	GC Bias32768	GC Bias65536	GC Bias131072	GC Bias262144	GC Bias524288	GC Bias1048576	GC Bias2097152	GC Bias4194304	GC Bias8388608	GC Bias16777216	GC Bias33554432	GC Bias67108864	GC Bias134217728	GC Bias268435456	GC Bias536870912	GC Bias1073741824	GC Bias2147483648	GC Bias4294967296	GC Bias8589934592	GC Bias17179869184	GC Bias34359738368	GC Bias68719476736	GC Bias137438953472	GC Bias274877906944	GC Bias549755813888	GC Bias1099511627776	GC Bias2199023255552	GC Bias4398046511104	GC Bias8796093022208	GC Bias17592186044416	GC Bias35184372088832	GC Bias70368744177664	GC Bias140737488355328	GC Bias281474976710656	GC Bias562949953421312	GC Bias1125899906842624	GC Bias2251799813685248	GC Bias4503599627370496	GC Bias9007199254740992	GC Bias18014398509481984	GC Bias36028797018963968	GC Bias72057594037927936	GC Bias144115188075855872	GC Bias288230376151711744	GC Bias576460752303423488	GC Bias1152921504606846976	GC Bias2305843009213693952	GC Bias4611686018427387904	GC Bias9223372036854775808	GC Bias18446744073709551616	GC Bias36893488147419103232	GC Bias73786976294838206464	GC Bias147573952589676412928	GC Bias295147905179352825856	GC Bias590295810358705651712	GC Bias1180591620717411303424	GC Bias2361183241434822606848	GC Bias4722366482869645213696	GC Bias94447329657392904271392	GC Bias188894659314785808542784	GC Bias37778931862957161708576	GC Bias75557863725914323417152	GC Bias151115727451828646842304	GC Bias302231454903657293684608	GC Bias604462909807314587369216	GC Bias12089258196146291754738304	GC Bias24178516392292583509476608	GC Bias48357032784585167018953216	GC Bias96714065569170334037906432	GC Bias193428131138340668077812864	GC Bias386856262276681336155625728	GC Bias773712524553362672311251456	GC Bias1547425049106725344622502912	GC Bias3094850098213450689245005824	GC Bias61897001964269013784899916512	GC Bias123794003928538027567799833024	GC Bias247588007857076055135599666048	GC Bias495176015714152110271199332096	GC Bias990352031428304220542398664192	GC Bias1980704062856608441084797328384	GC Bias3961408125713216882169594656768	GC Bias79228162514264337643391891133376	GC Bias1584563250285287552867837822672	GC Bias3169126500570575105735576454544	GC Bias6338253001141151011471128889088	GC Bias1267650600228230202274257779776	GC Bias25353012004564604045485155555936	GC Bias507060240091292080909703111119744	GC Bias10141204801825841618191406222239488	GC Bias20282409603651683236383812444479936	GC Bias405648192073033664727676248889584	GC Bias811296384146067329455352497779168	GC Bias1622592768292134658907048995558336	GC Bias3245185536584269317814097991116672	GC Bias6490371073168538635628195982233248	GC Bias129807421463717732712516391646664	GC Bias259614842927435454425032783293328	GC Bias519229685854870908850065566586656	GC Bias1038459371709401817600131133173312	GC Bias2076918743418803635200262266346624	GC Bias4153837486837607270400524532693248	GC Bias8307674973675214540801049065386496	GC Bias16615349947350428881602098130739904	GC Bias3323069989470085776320419626047888	GC Bias6646139978940171552640839252097776	GC Bias132922799578803431052816785041955552	GC Bias265845599157606862105633570083911104	GC Bias53169119831521372421127401638222208	GC Bias106338239663042744422244802566444416	GC Bias212676479326085488844489605132888896	GC Bias425352958652170977788899210273777792	GC Bias850705917304341955577798420547555584	GC Bias17014118346087039111555576401091111168	GC Bias340282366921740782231111152021822336	GC Bias680564733843481564462222304043644704	GC Bias13611294676869631288944446080870889408	GC Bias2722258935373926577888889216137177808	GC Bias5444517870747853155777778427273555168	GC Bias10889035741495703111555555685447111136	GC Bias2177807148299140622311111111270222272	GC Bias43556142965982812446222222254044444448	GC Bias8711228593196562489244444441080888896	GC Bias174224571839313249788888881761777777792	GC Bias34844914367864649957777
-----------	------	---------	-------------	----------------	---------	---------	----------	----------	----------	-----------	-----------	-----------	------------	------------	------------	-------------	-------------	-------------	-------------	--------------	--------------	--------------	---------------	---------------	---------------	----------------	----------------	----------------	----------------	-----------------	-----------------	-----------------	------------------	------------------	------------------	-------------------	-------------------	-------------------	-------------------	--------------------	--------------------	--------------------	---------------------	---------------------	---------------------	----------------------	----------------------	----------------------	----------------------	-----------------------	-----------------------	-----------------------	------------------------	------------------------	------------------------	-------------------------	-------------------------	-------------------------	-------------------------	--------------------------	--------------------------	--------------------------	---------------------------	---------------------------	---------------------------	----------------------------	----------------------------	----------------------------	----------------------------	-----------------------------	-----------------------------	-----------------------------	------------------------------	------------------------------	------------------------------	-------------------------------	-------------------------------	-------------------------------	--------------------------------	---------------------------------	--------------------------------	--------------------------------	---------------------------------	---------------------------------	---------------------------------	-----------------------------------	-----------------------------------	-----------------------------------	-----------------------------------	------------------------------------	------------------------------------	------------------------------------	-------------------------------------	-------------------------------------	--------------------------------------	---------------------------------------	---------------------------------------	---------------------------------------	---------------------------------------	--	--	---	--	--	--	--	---	--	--	--	--	--	---	---	---	--	--	--	---	---	---	---	--	---	---	---	---	--	---	---	---	---	---	---	---	---	--	--	---	--	---	--	--	--------------------------------

CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes BPF genes, which are homologues of the milk
CC protein butyrophilin (BTF), and can be used in the production of agonists
CC and antagonists of BTF function. The present sequence encodes BTF2. Also
CC described are: (1) a ROR γ gene which can be used to develop products
CC for the study, diagnosis and treatment of lupus and Sjogren's syndrome;
CC and (2) NPT3 and NPT4 genes which are homologues of a type 1 sodium
CC transport gene, and can similarly be used for hypophosphatemia.
XX
XX Sequence 3518 BP; 926 A; 831 C; 868 G; 893 T; 0 other:

Query Match	2.9%	Score 76.4	: DB 19	Length 3518
Best Local Similarity	57.4%	Pred. No. 2e-09		
Matches 218	Conservative 0	Mismatches 151	Indels 11	Gaps 4

[illegible]

RESULT	7
AC	AAC77707
AC	AAC77707 standard; cDNA; 2961 BP.
AC	AAC77707;
XX	
DT	08-FEB-2001 (first entry)
DE	
XX	Human cancer associated gene sequence SEQ ID NO:101.
XX	
KM	Human; cancer associated gene; cancer antigen; detection; cancer;
KM	diagnosis; cytostatic; proliferative; vulnereary; immunomodulator;
KM	antidiabetic; antilasthmatic; antihemematic; antiarthritic; antiviral;
KM	antiinflammatory; antithyroid; antiallergic; antibacterial; cardiact;
KM	dermatological; neuroprotective; thrombolytic; coagulant; noctropic;
KM	vasotropic; antiproliferic; antiangiogenic; gene therapy; inflammation;
KM	immune disorder; haematopoietic cell disorder; autoimmune disorder;
KM	allergic reaction; graft versus host disease; organ rejection;
KM	haemostatic; thrombolytic; cardiovascular disorder; infection;
KM	neurological disease; drug screening; ss.

OS	Homo sapiens.
XX	
PN	W0200055350-A1.
XX	
PD	21-SEP-2000.
XX	

PF 08-MAR-2000; 2000MO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX WPI; 2000-587533/55.
DR P-PSDB; AAB43498.
DR
XX Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
PT
XX Claim 1; Page 691-692; 2352PP; English.
XX

CC AAC77603 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytosstatic; proliferative; vulnerrary; immunomodulator;
CC antidiabetic; antilasthmatic; antihemematic; antialthritic;
CC antinflammatory; antihyproid; antiallergic; antibacterial; antiviral;
CC dermatologic; neuroprotective; cardiant; thrombolytic; coagulant;
CC neotropic; vasotropic; antispasitic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.

Sequence 2981 BP; 669 A; 807 C; 796 G; 706 T; 3 other;

Query Match	2.7%;	Score	69.2;	DB	21;	Length	2981;
Best Local Similarly	52.68;	Pred. No.	1.3e-07;				
Matches	224;	Conservative	0;	Mismatches	193;	Indels	9;
						Gaps	3;

QY	1396	tagatctcaagggttcatacaactcttgacccttcacagcccgcaagcgggagaccttgctctccg	1455
Db	1276	tatactcaagtgagacgtgjaacctctgagaccacagacggtctaacccaaag-ccttgatctctctc	1334
QY	1456	gaagacacgagaaagtctcaagttagacaccccgagacgaagaagagctctgacagacagccccctg	1515
Db	1335	gataatctcgtgcgaagtgcggttacaagtctaacctccaacagagactcgtctctgacaacacccgag	1394
QY	1516	cggcttcgaagcgccctcccgacggtctcgtggtcttccegggtcttcctccctccggcgccacccgc	1575
Db	1395	aggtctcaatctgcttccctctgctctggtcggtctccatctgcttatccgcgcggagacattat	1454
QY	1576	tggcagaagcttgaccctcgcaagcggcgacagccggcgacggtctgacagtgtaggggttgaccggagag	1635
Db	1455	tggagaggtctag-----agtgaggaagataaagccaagtggacaatctagtgtagtgtagaac	1508
QY	1636	gggtctgagagagacagagagagatactggagactcagcgcgcgaagagacggtctggtgcccgtat	1695
Db	1509	tcaagtgcagcaagaagaagtgtagagtaacctcaagccccccagaatagatctctggtcgagtgtct	1566
QY	1696	catctctgcacacaaagcagtgctctggtgcacagacacctccggcgaccagacctgtctcgctgag	1755
Db	1569	tctgtgfatgagaaagataatctggtctcttacctccccaatgactgcgc--taacctgtgc	1626
QY	1756	cggagatccgcgcgcaggtctgtagagctgcgcctctgagactacagagtcgggtgagctgtagacctcc	1815
Db	1627	ggacaccgcgtccacgcgggtctgtagaatctctcttgacacatgatagtctgtgtgtagtctctctct	1666

QY 1816 acaacy 1821
 Db 1687 acaacy 1692

RESULT 8
 AAX60263 standard; DNA; 43226 BP.
 ID AAX60263
 AC AAX60263;
 XX
 DT 12-AUG-1999 (first entry)
 DE Nucleic acid sequence from C121 gene of system B of chicken MHC.
 KW Resistance; tumour development; Marek disease tumour; chicken;
 KW system B; system Rfp-Y; poultry major histocompatibility complex; MHC;
 KW class II B-L gene; gene 17.5; gene 12.3; gene B-FIV; class I;
 KW genotype; selection; breeding; virus-induced tumour; C121 gene; ss.
 XX
 OS Gallus sp.
 XX FR2771422-A1.
 XX PD 28-MAY-1999.
 XX PF 21-NOV-1997; 97FR-0014669.
 XX PR 21-NOV-1997; 97FR-0014669.
 XX PA (CNRS) CNRS CENT NAT RECH SCI.
 XX DR WPI; 1999-329699/28.
 PT Chicken nucleic acid involved in controlling tumour susceptibility
 PS Claim 2; Fig 5; 49pp; French.
 XX
 CC The specification describes nucleic acid sequences from genes (or
 CC related regions) that encode proteins involved in controlling
 CC resistance or susceptibility to development of tumours (e.g. Marek
 CC disease tumours) in chickens. The nucleic acid sequences include
 CC sequences from gene of systems B or Rfp-Y of the poultry major
 CC histocompatibility complex (MHC), other than genes of class II B-L
 CC and genes 17.5, 12.3 or B-FIV of class I. The nucleic acid sequences
 CC are used to genotype poultry, particularly to select (for breeding)
 CC birds resistant to virus-induced tumours. The present sequence
 CC represents a nucleic acid sequence from a gene of system B.
 CC
 SO Sequence 43226 BP; 10365 A; 10522 C; 11244 G; 10981 T; 114 other;

Query Match 2.6%; Score 67.6; DB 20; Length 43226;
 Best Local Similarity 55.4%; Pred. No. 1.2e-06;
 Matches 237; Conservative 0; Mismatches 179; Indels 12; Gaps 5;

QY 1411 tcaactggaccctccagaccgacgagactggtctctcgtgaagaagaagtcga 1470
 Db 6583 tgactctggaccagagagcgccaccctcg-cctcgctctctccaagaagaccagagc 6641
 QY 1471 gtgaggtacaccgagcaagaagagcctcgcaagacagccctcgcttcgacgagctc 1530
 Db 6642 gtccgaatggagatacagctctcaggaatccccgagagccccgagcgcttcgaagcgcat 6701
 QY 1531 ccgagcggtctgggcttcctccggtctctcctccgggacgacgctggcaggttgacctg 1590
 Db 6702 cccgcgtgctgggtgttgaaaccttcacctctggagagcaactgctgggtg-----tg 6755
 QY 1591 cagctggagcgagcgagcgaggtgacaggtgggtggccgggggggggtgaagagagca 1650
 Db 6756 gatctacacaagaagcgactgacgcttgggggtcagcaagaggttccctgcacagga-a 6814
 QY 1651 gggagagatgggactcaagcgagcgagcgagcggtctggccgtgatcatctctgcaccaa 1710

Db 6815 aggagcgctcagcttcaaccttgatgaagcatctggcgctg--cagcaatgggggttc 6872
 QY 1711 cagtgctgggacgagacctcccggggacgacgactgtctccgtgagcagatccgcgacag 1770
 Db 6873 aagacagagcgctcctaccctccctccgacccaccatgaaaccttcacaggggttcccaaaag 6932
 QY 1771 gcgtgagagtcgacctgactacgagcgggcgaggtgtacctctccaacagcccgagacc 1830
 Db 6933 --atcgcatctctcgtgactacgaatggcgaggtggcgttttttgatgtgagaacc 6990
 QY 1831 caggggccc 1838
 Db 6991 aatggccc 6998

RESULT 9
 AAA78067
 ID AAA78067 standard; cDNA; 712 BP.
 XX
 AC AAA78067;
 XX
 DT 14-NOV-2000 (first entry)
 DE cDNA encoding human colon tumour polypeptide, SEQ ID NO:354.
 XX
 KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
 KW immunotherapy; diagnosis; progression; ss.
 XX
 OS Homo sapiens.
 XX WO200037643-A2.
 XX PD 29-JUN-2000.
 XX PF 23-DEC-1999; 99WO-US30909.
 XX PR 23-DEC-1998; 98US-0221298.
 XX PR 02-JUL-1999; 99US-0347496.
 XX PR 22-SEP-1999; 99US-0401064.
 XX PR 19-NOV-1999; 99US-0444242.
 XX PR 02-DEC-1999; 99US-0454150.
 XX PA (CORI-) CORIAX CORP.
 XX
 PT Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk J;
 PT Wang T, Yugu J;
 PT
 PS WPI; 2000-442671/38.
 XX
 CC Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
 CC portions of proteins which are associated with human colon tumours.
 CC The invention also specifically discloses 8 human colon tumour proteins
 CC (AA11897-B11904). The nucleic acids, the polypeptides they encode, and
 CC antigen presenting cells (APCs, preferably dendritic cells) expressing
 CC such polypeptides may be used in vaccines that target tumour cells,
 CC especially colon tumour cells, thereby inhibiting the development of
 CC cancer. T-cells specific for the polypeptide expressed by the APC are
 CC used to remove tumour cells from biological samples, especially blood or
 CC fractions thereof. The sample or the isolated T-cells specific for the
 CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
 CC CD8+ T-cells from a patient may be incubated with a polypeptide or
 CC nucleic acid of the invention, or an APC expressing such a polypeptide,
 CC to cause the proliferation of specific T-cells. The T-cells can be
 CC cloned and then administered back to the patient to inhibit cancer
 CC development. Nucleic acids encoding the polypeptides and antibodies
 CC against the polypeptides may be used to determine the expression level

PA (CNRS) CNRS CENT NAT RECH SCI.
 XX WPI; 1999-329699/28.
 DR
 XX
 PT Chicken nucleic acid involved in controlling tumour susceptibility
 XX
 PS Claim 2; Fig 5; 49pp; French.
 XX
 CC The specification describes nucleic acid sequences from genes (or
 CC related regions) that encode proteins involved in controlling
 CC resistance or susceptibility to development of tumours (e.g. Marek
 CC disease tumours) in chickens. The nucleic acid sequences include
 CC sequences from gene of systems B or Rfp-Y of the poultry major
 CC histocompatibility complex (MHC), other than genes of class II B-L
 CC and genes 17.5, 12.3 or B-FIV of class I. The nucleic acid sequences
 CC are used to genotype poultry, particularly to select (for breeding)
 CC birds resistant to virus-induced tumours. The present sequence
 CC represents a nucleic acid sequence from a gene of system B.
 XX
 SQ Sequence 43226 BP; 10365 A; 10522 C; 11244 G; 10981 T; 114 other;
 30
 Query Match 2.48; Score 63; DB 20; Length 43226;
 Best Local Similarity 53.8%; Pred. No. 1.7e-05;
 Matches 220; Conservative 0; Mismatches 180; Indels 9; Gaps 4;
 QY 1402 caggagtcatactactctgacccctcagaccgacgagcagctgtctctctggaagac 1461
 Db 36012 CAGCCAGCATACCTCGACCCCGACACCGCTCAC-CTGACCTCATTCTCTCGAAGAC 35954
 QY 1462 aggaagtcagtgagtaacacccgacagaagaagagctgcacagacccctgcgcttc 1521
 Db 35953 CGAAGAAGTGTGAACGTGGGGAAGACACAGACGCTTGATTAACCCGGAAGATTC 35894
 QY 1522 gacggtccctccgagctctgtggtctcccggtctctcccgagcgccacgctgcag 1581
 Db 35893 GCCTACTGGCCCTTTGTTTGGCCACCAAGCTTCTCTCGCGCGGCTGCTGGGAG 35834
 QY 1582 gttagcctcagctgtggagagcgagcgagcggtgtggtggtggcgagaggggtg 1641
 Db 35833 G-----TGCAGTGGGGGATGAAGGAGCTGGCCATGCTGTGGCCGGAATCCATC 35780
 QY 1642 aggaagacagaagaagatgagactcagcgacgagagcggtctggtgagctgatatc 1701
 Db 35779 CCTCGGA-AGGGTCACTACGCTCTGTCCCAAGGGGGGATTTGGGGGAGGAATG. 35721
 QY 1702 tgcacaaagcagctgtcgtggccagacactcccggtgcacgacgtctgcgtgagcgagat 1761
 Db 35720 GGGGGGACAGGTCCGGGCACTACCAACCCACAGGTGACCTGTAGCTGTGCGCTGGGT 35661
 QY 1762 ccgcgagcaggtgagagtgccctgagctacgagcggtgaggtgagc 1810
 Db 35660 GCCCAGGAGG-GTCAGCCTCCTTGGACTGAGCTGAGGAGCGGTGC 35613
 RESULT 12
 ID AAH72187 standard; CDNA: 376 BP.
 XX AAH72187;
 AC
 XX 19-SEP-2001 (first entry)
 DT
 XX Human cervical cancer marker nucleic acid 3461.
 DE
 XX
 KM Cervical cancer: cytostatic; pre-malignant condition; gene therapy; ss.
 OS Homo sapiens.
 XX
 PN WO200142467-A2.
 XX
 PD 14-JUN-2001.
 XX

PF 08-DEC-2000; 2000WO-US33312.
 XX
 PR 08-DEC-1999; 99US-0169681.
 PR 21-DEC-1999; 99US-0171350.
 PR 14-MAR-2000; 2000US-0189315.
 PR 12-MAY-2000; 2000US-0203791.
 PR 09-JUN-2000; 2000US-0210600.
 PR 21-JUL-2000; 2000US-0220114.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Deeds J, Berger A, Zhao X;
 DR WPI; 2001-375006/39.
 XX
 PT New isolated nucleic acid for diagnosing and treating cervical cancer
 PT and for assessing and detecting compounds for treating the cancer -
 XX
 PS Claim 1; Page 663-664; 1051pp; English.
 XX
 CC The invention relates to novel genes (AAH68727-AAH73383) associated with
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded
 CC polypeptides are useful: to assess if a patient is afflicted with
 CC cervical cancer or has a pre-malignant condition; to monitor the
 CC progression of cervical cancer of a premalignant condition in a patient;
 CC to select and/or assess the efficacy of a compound or therapy for
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be
 CC useful for gene therapy.
 XX
 SQ Sequence 376 BP; 84 A; 94 C; 107 G; 91 T; 0 other;
 50
 Query Match 2.38; Score 60.6; DB 22; Length 376;
 Best Local Similarity 60.78%; Pred. No. 7.9e-05;
 Matches 116; Conservative 0; Mismatches 74; Indels 1; Gaps 1;
 QY 1400 ttcaagggtcatcactctgcagaccctcagaccgacgagcagctgtctctcgagag 1459
 Db 64 tacagtgcagctgactctgcagaccgacgagcagcttaccacg-cctgactctctgata 122
 QY 1460 acaggaagtcagtgatcaccccgacagaagaagagctgcagacagacccctgcgct 1519
 Db 123 atctgcgcaagtggtgacagcttaccctcaacagagactgctgacacaccccgagagt 182
 QY 1520 tcgacggtccctccgaggttctggtctcccggtcttctctccggtgcacacgctgac 1579
 Db 183 tcaatctgttccctgtgtctgtggtctctcattcattcgtcgcggtgagacattatgg 242
 QY 1580 aggtgacctg 1590
 Db 243 aggtagaggtg 253
 RESULT 13
 ID AAH71127 standard; CDNA: 509 BP.
 XX AAH71127;
 AC
 XX 19-SEP-2001 (first entry)
 DT
 XX Human cervical cancer marker nucleic acid 2401.
 DE
 XX
 KM Cervical cancer: cytostatic; pre-malignant condition; gene therapy; ss.
 OS Homo sapiens.
 XX
 PN WO200142467-A2.
 XX
 PD 14-JUN-2001.
 XX
 PF 08-DEC-2000; 2000WO-US33312.
 XX

```
PR 08-DEC-1999: 99US-0169681.
PR 21-DEC-1999: 99US-0171350.
PR 14-MAR-2000: 2000US-0189315.
PR 12-MAY-2000: 2000US-0203791.
PR 09-JUN-2000: 2000US-0210600.
PR 21-JUL-2000: 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI: 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1, Page 504; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 509 BP; 103 A; 131 C; 149 G; 126 T; 0 other;

Query Match      2.3%; Score 60.6; DB 22; Length 509;
Best Local Similarity 60.7%; Pred. No. 9.1e-06;
Matches 116; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 1400 ttgaagggtacatcactctgacccctcagccgacgcgagacgtgtctctcggaag 1459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 94 taacagtgagcgtgactctctgacccagacagcagcgtactccag-cctgactctctcgata 152
QY 1460 acaggaagtcagtgaagtacacccgcagagaagaagcctgcagacagccctcgct 1519
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 153 atctgcgcagcgtgcgttacagttaccctccacagacagcctgctgacaaaccgagagt 212
QY 1520 tcgacggcctcccgcggttcctccggcttctcctccggcgccacgcgtgcg 1579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 213 tcaatctgttccctgtctctgtggtctccatgctcctcgcgcggagacattatggg 272
QY 1580 aggttgacctg 1590
   ||| ||| |||
DB 273 aggtagaagtg 283

RESULT 14
AAH72832
ID AAH72832 standard; cDNA; 750 BP.
XX
AC AAH72832;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 4106.
XX
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200142467-A2.
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999: 99US-0169681.
PR 21-DEC-1999: 99US-0171350.
```

```
PR 14-MAR-2000; 2000US-0189315.
PR 12-MAY-2000; 2000US-0203791.
PR 09-JUN-2000; 2000US-0210600.
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI: 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1, Page 876-877; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 750 BP; 191 A; 167 C; 196 G; 192 T; 4 other;

Query Match      2.3%; Score 60.6; DB 22; Length 750;
Best Local Similarity 60.7%; Pred. No. 1.1e-05;
Matches 116; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 1400 ttgaagggtacatcactctgacccctcagccgacgcgagacgtgtctctcggaag 1459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 438 tgacgtgcagcgtgactctgacccagacagcagcgtactccag-cctgactctctcgata 496
QY 1460 acaggaagtcagtgaagtacacccgcagagaagaagcctgcagacagccctcgct 1519
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 497 atctgcgcagcgtgcgttacagttaccctccacagacagcctgctgacaaaccgagagt 556
QY 1520 tcgacggcctcccgcggttcctccggcttctcctccggcgccacgcgtgcg 1579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 557 tcaatctgttccctgtctctgtggtctccatgctcctcgcgcggagacattatggg 616
QY 1580 aggttgacctg 1590
   ||| ||| |||
DB 617 aggtagaagtg 627

RESULT 15
AAV05159
ID AAV05159 standard; DNA; 14180 BP.
XX
AC AAV05159;
XX
DT 20-JUL-1998 (first entry)
XX
DE Mouse butyrophilin gene.
XX
DE Butyrophilin: b1n gene; promoter; mouse; transgenic animal;
KW mammary gland-specific gene expression; breast cancer; diagnosis;
KW ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT Promoter 1..4693
   /*tag= a
FT TATA_signal 4645..4651
   /*tag= b
FT prim_transcript 4611
   /*tag= c
```

```
FT /evidence= EXPERIMENTAL
FT /standard_name= transcription start site
FT /note= "most frequently used site"
FT misc_signal
FT 4645..4651
FT /*tag= d
FT /note= "AT-rich, TATAA-like element"
FT misc_signal
FT 4583..4588
FT /*tag= e
FT /note= "AT-rich, TATAA-like element"
FT misc_signal
FT 4597..4603
FT /*tag= f
FT /note= "CCAAT-like element"
FT misc_signal
FT 4546..4551
FT /*tag= g
FT /note= "CCAAT-like element"
FT misc_signal
FT 4533..4537
FT /*tag= h
FT /note= "CCAAT-like element"
FT prim_transcript
FT 4675
FT /*tag= i
FT /evidence= EXPERIMENTAL
FT /standard_name= transcription start site
FT prim_transcript
FT 4694
FT /*tag= j
FT /evidence= EXPERIMENTAL
FT /standard_name= transcription start site
FT polyA_signal
FT 13091..13096
FT /*tag= k
FT /evidence= EXPERIMENTAL
FT prim_transcript
FT 13097..13199
FT /*tag= l
FT /evidence= EXPERIMENTAL
FT /standard_name= 3' end of transcript
FT CDS
FT 4923..11397
FT /*tag= m
FT /evidence= EXPERIMENTAL
FT /note= "contains introns"
FT sig_peptide
FT 4923..5001
FT /*tag= n
FT 4650..4651
FT /*tag= o
FT /evidence= EXPERIMENTAL
FT /standard_name= translational initiation codon
FT CDS
FT 4743..4745
FT /*tag= p
FT /evidence= EXPERIMENTAL
FT /standard_name= translational initiation codon
FT CDS
FT 4765..4767
FT /*tag= q
FT /evidence= EXPERIMENTAL
FT /standard_name= translational initiation codon
FT CDS
FT 4776..4778
FT /*tag= r
FT /evidence= EXPERIMENTAL
FT /standard_name= translational initiation codon
FT exon
FT 4611..5001
FT /*tag= s
FT /codon_start= 4923..4925
FT /number= 1
FT intron
FT 5002..5520
FT /*tag= t
FT /evidence= EXPERIMENTAL
FT /number= A
FT exon
FT 5521..5871
FT /*tag= u
FT /number= 2
FT intron
FT 5872..8332
FT /*tag= v
FT /evidence= EXPERIMENTAL
FT /number= B
FT exon
FT 8333..8614
FT /*tag= w
FT /number= 3
```

```
FT intron
FT 8615..9485
FT /*tag= x
FT /evidence= EXPERIMENTAL
FT /number= C
FT exon
FT 9486..9635
FT /*tag= y
FT /number= 4
FT intron
FT 9636..10206
FT /*tag= z
FT /evidence= EXPERIMENTAL
FT /number= D
FT exon
FT 10207..10227
FT /*tag= aa
FT /number= 5
FT intron
FT 10228..10320
FT /*tag= ab
FT /evidence= EXPERIMENTAL
FT /number= E
FT exon
FT 10321..10347
FT /*tag= ac
FT /number= 6
FT intron
FT 10348..10738
FT /*tag= ad
FT /evidence= EXPERIMENTAL
FT /number= F
FT exon
FT 10738..13199
FT /*tag= ae
FT /number= 7
FT repeat_unit
FT 4807..4814
FT /*tag= af
FT /rpl_type= INVERTED
FT repeat_unit
FT 12556..12563
FT /*tag= ag
FT /rpl_type= INVERTED
FT protein_bind
FT 2948..2960
FT /*tag= ah
FT /note= "NFI"
FT protein_bind
FT complement (2971..2979)
FT /*tag= ai
FT /note= "gamma-IRE"
FT protein_bind
FT complement (3010..3015)
FT /*tag= aj
FT /note= "myb"
FT protein_bind
FT complement (3030..3036)
FT /*tag= ak
FT /note= "YY1"
FT protein_bind
FT complement (3115..3123)
FT /*tag= al
FT /note= "gamma-IRE"
FT protein_bind
FT 3129..3135
FT /*tag= am
FT /note= "YY1"
FT protein_bind
FT 3132..3146
FT /*tag= an
FT /note= "NFI"
FT protein_bind
FT 3138..3146
FT /*tag= ao
FT /note= "STAT"
FT protein_bind
FT complement (3145..3150)
FT /*tag= ap
FT /note= "GMCSF"
FT protein_bind
FT 3181..3189
FT /*tag= aq
FT /note= "C/EBP"
FT protein_bind
FT complement (3187..3197)
FT /*tag= ar
FT /note= "C/EBP"
FT protein_bind
FT 3190..3195
FT /*tag= as
FT /note= "PU.1"
FT protein_bind
FT complement (3194..3200)
FT /*tag= at
FT /note= "alpha-IFN"
```

```
FT protein_bind 3195..3203
FT /tag= au
FT /note= "GMCSF"
FT protein_bind 3253..3259
FT /tag= av
FT /note= "gamma-IRE"
FT protein_bind 3370..3378
FT /tag= aw
FT /note= "GRE"
FT protein_bind complement (3397..3406)
FT /tag= ax
FT /note= "GRE"
FT protein_bind complement (3399..3403)
FT /tag= ay
FT /note= "gamma-IRE"
FT protein_bind 3419..3427
FT /tag= az
FT /note= "STAR"
FT protein_bind 3439..3445
FT /tag= ba
FT /note= "gamma-IRE"
FT protein_bind 3450..3454
FT /tag= bb
FT /note= "GMCSF"
FT protein_bind 3456..3462
FT /tag= bc
FT /note= "STAR"
FT protein_bind 3457..3463
FT /tag= bd
FT /note= "C/EBP"
FT protein_bind complement (3464..3468)
FT /tag= be
```

```
Query Match 2.3%; Score 60.6; DB 19; Length 14180;
Best Local Similarity 54.6%; Pred. No. 4.3e-05;
Matches 230; Conservative 0; Mismatches 179; Indels 12; Gaps 5;

QY 1400 ttcagggtatcaactctgacccctcagaccgcgacgacgtgtctctcggaag 1459
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10728 tccagttgacgtgactctgatacagacagaccacc-ccacctctctgtatgaag 10786

QY 1460 acagggaagtcagtgaggtacaccgcgacgaagaagacgtgcgacagccctgcgc 1519
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10787 attcaagtcagtcgtcgtatggaagattcacgtccgacccgtgatagaccagaagat 10846

QY 1520 tcgacgacctcccgcggtctctgggtctcccggtctctcccgcgccacgcgtgc 1579
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10847 ttgactccttgccctgtgtgttgccgtgagaccttacttcaggagagatactctgg 10906

QY 1580 aggttgacctgcagctgcgcagcgcggtgcgtgcacggtggtgggtggtccgggagggg 1639
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 10907 aggttgg-----aggtggagatatagactgactggccattggtgtgtagggagaaatg 10960

QY 1640 tggaggaggacaggagagatggagctcagcgccgagagacggtctgtggccgtgatcatc 1699
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 10961 tgggtgagaagaagggtttgacccca-tgactccgataatgggtcttgggtctggaagttg 11019

QY 1700 tcgtccaccaagcagtgctggcgccagcacctcccggtccacgacctgtccgtgagcgag 1759
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 11020 tatgga--aatgggtactgtggccctcaccactcagagacc--tctctcgattagcagg 11075

QY 1760 atcccgcgacagcgctgaagtcgacctggaactcgaagcggtggcgaagtgacctccacaa 1819
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 11076 gccccctcgagagttggttttctctgactatgacgcagagacatttccttctacaa 11135

QY 1820 c 1820

Db 11136 c 11136
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 13:34:39 ; Search time 8456.85 Seconds

(Without alignments)
5091.451 Million cell updates/sec

Title: US-09-867-034-5

Perfect score: 2610
Sequence: 1 gatgtgagcagcctcagag.....ccgagatggtcgcagagagt 2610

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1765.4	67.6	2255	9	HSU34249	U34249 Human putat
2	1301.4	49.9	1734	9	AF220132	AF220132 Homo sapi
3	971.2	37.2	47777	9	AC004189	AC004189 Homo sapi
4	971.2	37.2	100000	9	AP000517	AP000517 Homo sapi
5	971.2	37.2	162383	9	AB023054	AB023054 Homo sapi
6	971.2	37.2	162383	9	AB023055	AB023055 Homo sapi
7	714.4	27.4	17284	9	HSR1NGG	Y07829 Homo sapien
8	332.4	12.7	152211	4	SSC251829	AF220122 Homo sapi
9	271.4	10.4	318	9	HS73B8R	AF220121 Mus muscu
10	246.8	9.5	426	9	AF220133	AF134811 Mus muscu
11	198.8	7.6	158414	10	AC005960	AF220123 Rattus no
12	191.6	7.3	2002	9	AF220122	AF220122 Homo sapi
13	170.4	6.5	2269	10	AF220121	AF134811 Mus muscu
14	165.6	6.3	2214	10	AF134811	AF134811 Mus muscu
15	160.6	6.2	54674	2	RN499P20	AL603720 Rattus no
16	137.2	5.3	3031	9	AF220123	AF220123 Homo sapi
17	117.6	4.5	2151	9	AF396651	AF396651 Homo sapi
18	116	4.4	3595	9	HSU09825	U09825 Human acid
19	105.8	4.1	47777	9	AC004189	AC004189 Homo sapi
20	105.8	4.1	100000	9	AP000517	AP000517 Homo sapi
21	105.8	4.1	152211	4	SSC251829	AF220123 Rattus no
22	105.8	4.1	162383	9	AB023054	AB023054 Homo sapi
23	105.8	4.1	162383	9	AB023055	AB023055 Homo sapi
24	102.6	3.9	17284	9	HSR1NGG	Y07829 Homo sapien
25	98	3.8	281	9	HS73B8R	Z62789 H.sapiens C
26	97.4	3.7	1888	9	AF396652	AF396652 Homo sapi
27	97.4	3.7	1949	9	AF396654	AF396654 Homo sapi
28	97.4	3.7	1956	9	AF396653	AF396653 Homo sapi
29	97.4	3.7	2328	9	AF396655	AF396655 Homo sapi
30	92.8	3.6	789	9	HST000221	AL360164 Homo sapi
31	90.2	3.5	120524	9	AC008443	AC008443 Homo sapi
32	89.6	3.4	236913	2	AL161615	AL161615 Homo sapi
33	84.4	3.2	2296	9	HST000220	AL360163 Homo sapi
34	82.8	3.2	174612	2	AC023889	AC023889 Homo sapi
35	81.4	3.1	1821	9	BC011629	BC011629 Homo sapi
36	79.4	3.0	183399	2	AC008620	AC008620 Homo sapi
37	76.6	2.9	158414	10	AC005960	AC005960 Mus muscu
38	76.4	2.9	3502	6	AR036568	AR036568 Sequence
39	76.4	2.9	3578	9	HSU90550	U09550 Human butyr
40	76.4	2.9	39412	6	A95300	A95300 Sequence 31
41	76.4	2.9	170001	9	HS45P21	AL021917 Human DNA
42	76.2	2.9	134331	2	RN462P8	AL603727 Rattus no
43	76	2.9	164766	2	AC026657	AC026657 Homo sapi
44	75.4	2.9	54674	2	RN499P20	AL603720 Rattus no
45	69.2	2.7	1782	9	HUMREPA	J03407 Human rfp t

ALIGNMENTS

RESULT	1	LOCUS	HSU34249	2255 bp	mRNA	PRI	12-JAN-1999
DEFINITION			Human putative zinc finger protein (ZNF87) mRNA, complete cds.				
ACCESSION			U34249				
VERSION			U34249.1	GI:4096653			
KEYWORDS							
SOURCE			human.				
ORGANISM			Homo sapiens				
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS			Goel,V.L., Parimoo,S., Capossela,A., Chu,T.W. and Gruen,J.R.				
TITLE			Isolation of novel non-HLA gene fragments from the hemochromatosis region (6p21.3) by cDNA hybridization selection				
JOURNAL			Am. J. Hum. Genet. 54 (2), 244-251 (1994)				
MEDLINE			94136483				
REFERENCE			2 (bases 1 to 2255)				
AUTHORS			Goel,V.L., Kuida,S., Chu,T.W., Bowlds,C. and Gruen,J.R.				
TITLE			A cluster of genes proximal to HLA-A encode putative zinc finger				

proteins
 JOURNAL Unpublished (1995)
 REFERENCE 3 (bases 1 to 2255)
 AUTHORS Goel, V. L.
 TITLE Direct Submission
 JOURNAL Submitted (17-AUG-1995) Vita Goel, Pediatrics, Yale University
 School of Medicine 333 Cedar Street, New Haven, CT 06510-8064, USA
 FEATURES
 Location/Qualifiers
 source 1..2255
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6p21.3; 150 kb centromeric of HLA-A"
 gene 1..2255
 /gene="ZNF87"
 1..493
 /gene="ZNF87"
 repeat_region 329..380
 /rpt_type=tandem
 /rpt_unit=ct
 494..1891
 CDS
 /gene="ZNF87"
 /note="The predicted amino acid sequence of the ZNF87 cDNA contains two zinc finger motifs at the N-terminus. The first is a C3HC4 and the second is a CHC32 zinc finger motif. In the middle is a sequence containing heptads of hydrophobic amino acids which theoretically can form a coiled coil structure reminiscent of a leucine zipper. This protein is a potential zinc finger transcription factor because of these features; putative zinc finger protein"
 /codon_start=1
 /product="zinc finger protein"
 /protein_id="ABD03787.1"
 /db_xref="GI:4096654"
 /translation="MPATPSLKVHVEHLPACTLCAGPLEDAVTPCGHTFRCLEPLALS
 OMAOSSGKILPLCQEEQAEPAAPVLPGLPGETYCEHGEKLYFECENDAEFLC
 VECRBPQHAFTVGLDEAIQPYRDLRSLEALSTEDIEDVKOCDDKLOVLT
 QIESKHQVETAFERLOQLERQCLRLRLRELQKIKERDEYTKVSEVRLTA
 OVKELEKCOQASLELDQVRVNSKREKTPSPALSPDLVKIRDPKRLTLEPE
 MRMFSENLAHLIEDSGVITLDPTQASLSLSEBRKSVRTYRROKSLPDLRPPDG
 LPAVAGFPGSSGRHRMVDLDLGGCGCTVAGGVRKREKMSADQWYAVIIS
 HQCWASTSPGDLPLSELPRGRVALDTEAGQVILHNAOTDEPLTFITFASFGKVP
 FAVWKKSGSLTLKG"

BASE COUNT 482 a 640 c 686 g 447 t
 ORIGIN
 3'UTR
 /gene="ZNF87"
 1892..2255

Query Match 67.6%; Score 1765.4; DB 9; Length 2255;
 Best Local Similarity 93.0%; Pred. No. 0;
 Matches 2142; Conservative 0; Mismatches 111; Indels 50; Gaps 26;

QY 1 gatgtggcagcgcctcaagagcagaattatagctcccaactgctcaatctacagga 60
 |||||
 DB 1 GATGTGGCAGCGCTCAGACCGAAGATTATGCTCCCACTGCTCAATCTACAGGAA 60
 QY 61 gctctgtcccccagcttcccccagcactgtgtcttaagattccaagaaaccatccc 120
 |||||
 DB 61 GCTTCTGCTCCCCAGTTCTCCAGCCACTGTGTCTACAGATTCCAGAAACCCATCCC 120
 QY 121 cctgtgacctcagagtgctctctctccacctagagcagcagaagcagagta 180
 |||||
 DB 121 CCGTGACCTCAGAGGTGTGTCTCTCCACCTAGGGAGCAGAGACCCAGAGTA 180
 QY 181 agaactggtctacttgccgcgcacatctctgggtaattcagagcgccttggaat 240
 |||||
 DB 181 AGAAGTGGCTTACTTGCCCGCCACATGGAAATTCGTGGTAATTCGAGACCCCTGGAAAT 240
 QY 241 tggaccacactccgctgataggtgtgtggcagaggttctaaggaacaaagaggcgagca 300
 |||||
 DB 241 TGGACCACTCCGCTGATAGTGTGTGGCGAGGCTCTAGGGAACACAGAGGCGAGCCA 300

QY 301 ggtggtccctgtgtgtgcatctctgtctctctctctctctctctctctctctct 360
 |||||
 DB 301 GGTGGCTTCCCTGTGCTGGCATTTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
 QY 361 ctctctctctctctctctcaagccttgaagccggttccctccctcgctcaagtgtg 420
 |||||
 DB 361 CTCTCTCTCTCTCTCTCTCAAGCTTGAAGCCGTTTCCCTCTGCG-----GATTCATGTA 415
 QY 421 aagtgacctcagatttaaggaaggaactcgtgtgtgtgtgtgtgtgtgtgtgtgt 480
 |||||
 DB 421 AAGTGACCTGATTTACAGGAAGGAAAGGAACTGCGTGGCTGAGAGACCGAGATGACGG 475
 QY 481 gctggggaagcagcctgtatgtcccgcaaccccgctccctgtgaaggggtgtgtccatgac 540
 |||||
 DB 481 GCTGGGGAAGCGACCGCTGATGCCCGGACACCCCTGCC-----TGAAGGTGCTCATGACG 530
 QY 496 gctgtgacctgtaccctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
 |||||
 DB 531 TGCTGCTCTGTACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 590
 QY 541 tgcctgtgacctgtaccctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
 |||||
 DB 591 ACACCTTTGCGCGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 649
 QY 601 acaactctgtgacctgtaccctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 660
 |||||
 DB 661 AAGATCTGCTCTGTGCGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
 QY 650 AAGATCTGCTGTCGCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 709
 QY 721 gtgtccctgtgacctgt 780
 |||||
 DB 710 GTGCCCCCTGGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 768
 QY 781 ctctctgtgagaagcagatgt 840
 |||||
 DB 769 CTTC-TGGGGAAGCATGCGAGTTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 827
 QY 841 aagcagcacacgt 900
 |||||
 DB 828 AGCGCAGACCGGCGGTTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 887
 QY 901 gtgtgactgtgaagctgt 960
 |||||
 DB 888 GTGTGACTGTGAAGCTGT 946
 QY 961 gaccagaagcttcaagctgt 1020
 |||||
 DB 947 GACCAAGCTTCAAGT 1004
 QY 1021 cacaagctccttgagaagctgt 1080
 |||||
 DB 1005 CA---CCTTTTGAAGAGGTGCG-AGCAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGT 1060
 QY 1081 cagagctgt 1140
 |||||
 DB 1061 AGGCTGAGGAGCT-----GGAGCAGCAGATTGTGAAGAGAGGAGTAAATATACACA 1114
 QY 1141 aaggtctctgtgaagctcaaccggtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1200
 |||||
 DB 1115 AAGGTCTCTGAGGAATCAACCGGCTTGGAG-CCGAGGTCAAGAGACT-GGAGAGAAAT 1172
 QY 1201 gtcaagcagcagcaatgt 1260
 |||||
 DB 1173 GTGAGCAGCAGCAAGTGAAGCTTCTCAAGAGTGTCAAGAGTCAAA-CCAGAGCAGGTGTAG 1231
 QY 1261 atgaagaacttgt 1320
 |||||
 DB 1232 ATGAAAGACTTTGT 1290
 QY 1321 ttccacaggaataactaaccttccacagagatgtatgtgaatgtgtgtgtgtgtgtgtgt 1380
 |||||
 DB 1291 TTTCACAGGAATATCTACACCTCCCAAGATGATGAGAGATGTTCTC-AGAAACTTGG 1349

QY	1381	cgatcaactctgaaatagattcaagggttcatactacttgcagccccaacccgcaagccgga	1440
Db	1350	CGCATCATCTGGAAATGATTTCAAGGGGTCATCACTCTGGACCTCTCAACCGCCAGCCGGGA	1409
QY	1441	gaactgtgtctctctcggaagacaggaaatcgaatgaggtatcaccccgcaagaagaatgcctg	1500
Db	1410	G-CCGTGGTCTCTGGAAAGACAGGAAGTCAGTGAAGTACACCCGGCAGAAAGAACGCTTG	1468
QY	1501	ccagacagccccctgtcgcttcacagcgccctccggcgagtlcttgggtctccggagctctcc	1560
Db	1469	CCAGACAGCCCCCTCGCGCTTTCAGAGCGCTCCCGCGGTTCTTGCGGCTTCCCGGCTTCTCC	1528
QY	1561	tcggcgcgcaaccccgctgcaggttgacctgagcttggcgagctgcagcgctgcagtgatg	1620
Db	1529	TCCGGGGCCCAACCCCTGGCAAGTTGACCTGAGCTGGAGCGACGGCGCGGCTGCACCGTG	1588
QY	1621	gggggtgcgggggggggtgtaggaaggacaaggagagatgtgagctcaagcgccagagacg	1680
Db	1589	GGGGTGGCGGGGAGAGGGGCTTAGAGAGA-AGGAGAGATGGGACTCAGCCCGCAGAGACGG	1647
QY	1681	cgctctgggcgtgatalcatctctctgcaccaagcagtgctctgggcagcaacctcccgcgacc	1740
Db	1648	CGTCTGGGCGCTGATCATATCTC--GCACACAGAGTGTGGGCGACAGCACTCCCGCGGCACC	1705
QY	1741	gaactgtctcgtctgagcagatcccgcgcaagcgtgtgaatgctgccctgagtaagagcg	1800
Db	1706	GACCTG-CCGCTGAGCGAGATCCCGCGC-GCGGTGAAGATGTCCTTGACTACAGAGCGG	1763
QY	1801	ggcagggtgaacctccacaagcccaagagcccaaggagccatcccttgaactcatgtgctc	1860
Db	1764	GGCAGGTACCTCTCACACGCCCCAGA--CCAGAGGCCATCTTCACTTCACTGCGCTC	1821
QY	1861	ttctctccggccaaggtcttctcctgtctcttgcgcgccttgcacacaaaggctctgcgctt	1920
Db	1822	TTT--CTTCCGGCAAAAGTCTTCCCTTCTTCCGTCTGG--AAAAAAGTTCCTGCGCTT	1876
QY	1921	agcgtgcacagggggaatgtagggcgcgcaaggcgcgagcgagacgagcgtctcc	1980
Db	1877	ACGCTGA--AAGCTGAAGTGGGGCGCGGAAAGGGCGGCAAGACCGAGACGGCGGCTCTCC	1935
QY	1981	ggagctcagagtcgcgcctctgacagtgltgcgcccgggggtctccctgtgccggtgag	2040
Db	1936	GGGATCAAGCTCCGCCCTTGCGCAAGTGTGCGGCCCGGGGGCTCTCTTGCCCGGCTAGG	1995
QY	2041	cgaaagaaacacggagagactgagctcgaacagcggtgtgtttacttaatttacttag	2100
Db	1996	CGAAGAA--CAGGGGACTTGAAGTCTCGAACAAGGGGTGTTTACTTTATTATCTTAG	2053
QY	2101	gcctcagctctcctgtacgctctcgaagctccctctgtacgctcttgcgctctctgcactca	2160
Db	2054	GCCCTCAGCTCCCTGACGCTCTGAGCCCTCCTGTGACGCTCTGCGCTTCTGTGACCTCA	2113
QY	2161	gaatgcagaacacaagcagcttgcgtctgtgccttaaggcaaacagccaacttaagaaaccg	2220
Db	2114	GAGTGCAGAAACCAAGAGAGGCTTGGGCTGTCCCTAGGGCAACACCCAACTTAGAGGCTAG	2173
QY	2221	ccggcctctgcggggaagaactcaagaaagagagacactcaaaatgtaatgtttaaactgttt	2280
Db	2174	CGGGCTTTCGGGGGAAAAA--AAAGAAAAAGACATCTTAATAATTAATGTTTAAACGTGTTT	2230
QY	2281	caagataatactcttgcggaaaaa 2303	
Db	2231	CAAAATAAAAAATAAAAAAAA 2253	
RESULT	2		
AF220132			
LOCUS	AF220132	1734 bp	mRNA
DEFINITION	Homo sapiens tripartite motif protein TRIM15 alpha (TRIM15) mRNA,		
ACCESSION	AF220132	complete cds, alternatively spliced.	
VERSION	AF220132.1	GI:12407432	
KEYWORDS			

Source	Organism	Human
REFERENCE	Human	Human
AUTHORS	Raymond A., Meroni G., Fantozzi A., Merla G., Cairo S., Luzzi L., Rijnance L.D., Zanatta E., Messali S., Calarca S., Guffanti A., Minucci S., Pellicci P.G. and Ballabio A.	
TITLE	The tripartite motif family identifies cell compartments	
JOURNAL	EMBO J. 20 (9), 2140-2151 (2001)	
PUBMED	11331580	
REFERENCE	2 (bases 1 to 1734)	
AUTHORS	Raymond A. and Meroni G.	
TITLE	Direct Submission	
JOURNAL	Submitted (30-DEC-1999) TIGEM, Via Olgettina 58, Milan, MI 20132, Italy	
FEATURES	Location/Qualifiers	
Source	1..1734	
gene	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	1..1734	
	/gene="TRIM15"	
CDS	1..1398	
	/gene="TRIM15"	
	/note="alternatively spliced"	
	/codon_start=1	
	/product="Tripartite motif protein TRIM15 alpha"	
	/protein_id="AA653505.1"	
	/db_xref="GI:12407433"	
	/translation="MPATPSKIVHELPACTLCAPLIEDAVTTCGTHFCRLCLPALISOMGSSGKILLCPLOEEDAEFPMAPVPLGSEYCEHGERIYFENDAEFLCVFCEGPTQHOATVGLFDEALQYRDRIRLSLELSSTEREIEIVKCOEDKLOLVLTQIESKHVETAFERLOOEQRCILARLELOOIMKRDYITKVSSEYRLGAOVKLEEKCOOPASELQEDVYVNSCRMLKFEVSPALISPLVKKIDPHKLTLPDMRFESEINLAHLIEDSGVITLDQTSRSYLSIEDSKSVYTRQKSLSPSLRPFQGLPVLAVGPPGSSGSHRNQVNDQLQEDGGCTVGAEGEVRKGEGLSDEADVMVITIS HQDCVASTSPGTDLPLEISIPRGVVALDYENAGVYTLNNAQDEPFIFFTASFSGVKPPPEFAVWKKSGCSLTLLG"	
BASE COUNT	369 a 497 c 547 g 321 t	
ORIGIN		
Query Match	49.9%; Score 1301.4; DB 9; Length 1734;	
Best Local Similarity	93.1%; Pred. No. 7.4e-285;	
Matches 1630; Conservative	0; Mismatches 81; Indels 40; Gaps 24;	
527	ggtgtgcataagctgctgctgctgtaccctctgtgcggcgctgtagagatgcggttac	586
24	ggtgtgcataagctgctgctgctgtaccctctgtgcggcgctgtagagatgcggttac	83
587	catctcgtgtgaacacacctctgcgcgctctgcctcccgcgctcccaatgtagggc	646
84	catctcgtgtgaacacacctctgcgcgctctgcctcccgcgctcccaatgtagggc	143
647	ccaatcctcgttggacaagaatcctgtcctctgcgcgctctgcgaaggagatgacagacaaga	706
144	ccaatcctcgttggacaagaatcctgtcctctgcgcgctctgcgaaggagatgacagacaaga	202
707	ctcccaatgagccctgtgcccctggcgccgtggagagataacttacttcgagaggacagcgc	766
203	ctcccaatgagccctgtgcccctggcgccgtggagagataacttacttcgagaggacagcgc	261
767	gagaagaactacttctctcttgcgagaaagatgagcgaagttccctctgtgttcctgagaga	826
262	gagaagaactacttctctcttgcgagaaagatgagcgaagttccctctgtgttcctgagaga	320
827	gggtcccaagcaccagcagcagccgttgagggttccctgagcagaggccattcagccctacgc	886
321	gggtcccaagcaccagcagcagccgttgagggttccctgagcagaggccattcagccctacgc	380
887	ggatcgtctcagagatgcacttggaagctcttgagcagcagagagatgagattgttagatg	946
381	ggatcgtctcagagatgcacttggaagctcttgagcagcagagagatgagattgttagatg	439

[illegible]

Qy	2027	gtgcgcgcgtbaagcgagagaacaacgaggaacttgagctctgaaacagcgattgttttac	20866
Db	1489	gtgcgcgcgtbaagcgagagaaca--caggggacgttgagcttcacacacggcgtttgttttac	15466
Qy	2087	tttatctatctcttaagccctcagctccctcctgaagtcagctccctctgtgaagctctgcgc	21466
Db	1547	ttttatttatctttaaagccctcagctccctcctgaagtcagctccctcctgtgaagcctcttcgccc	16066
Qy	2147	ttctctgcacctgaagatgtagaagacaacagagcgtctgagctgtgctgtgctctaaaggaacagcc	22066
Db	1607	tttctctgcacctgaagatgtagaagacaacacagcggcttgcgctgtgtccttaagggcacacagcc	16666
Qy	2207	aacctagaagacccgcgcgcctcttcggtggaataaactaaagaagagacatctaaatgttaa	22666
Db	1667	aaacttagagagccacacggcgcttctgcgggaaaaa---aaagaaaaaacacatctttaaataaaaa	17236
Qy	2267	tgatttaagctg	2277
Db	1724	tgatttaagctg	1734

LOCUS	AC004189/c	RESULT 3	PI	17-JUL-1998
DEFINITION	Homo sapiens clone UMC:Y18c247 from 6p21, complete sequence.			
ACCESSION	AC004189			
VERSION	AC004189.1	GI:3327409		
KEYWORDS	HTG.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 47777) Janer,M.M., Guillaudeau,T., Vu,O., Kutayavln,T., Harter,H. and Geraghty,D.E.			
TITLE	Large scale sequence analysis of the human MHC class I region			
JOURNAL	Unpublished (1998)			
REMARK	Fred Hutchinson Cancer Research Center The Clinical Research Division 1100 Fairview Ave. N., P.O. Box 19024 Seattle, WA 98109-1024			
REFERENCE	2 (bases 1 to 47777) Geraghty,D.E. and Olson,M.V.			
AUTHORS	Direct Submission			
TITLE	Submitted (23-FEB-1998)			
JOURNAL	Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA			
REFERENCE	3 (bases 1 to 47777) Geraghty,D.E. and Olson,M.V.			
AUTHORS	Direct Submission			
TITLE	Submitted (17-JUL-1998)			
JOURNAL	Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, Seattle, WA 98195, USA			
REMARK	Contact: Daniel E. Geraghty (geraghty@ehc.org) On Jul 17, 1998 this sequence version replaced gl:2905871.			
COMMENT	Overlapping Sequences: 5 : UMGc:y17c054 (genbank Accession: AC004186) 3 : UMGc:y2c224 (genbank Accession: AC004202)			

Sequence Quality Assessment: This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage:	88.8
DS or two chemistry coverage:	99.0
Single stranded regions:	2

Sequence Validation:

This sequence has been validated by Multiple Complete Digest

OY	1823	ccagagcccaaggccccctccctcaacctcaacttgacctctctctccgagcaagctctcc	1882
Db	15728	CCAGA--CCCAAGAGACCATCTTCACTTCACTGCTCTTT--CTCCGGCAAACTCTTCC	15673
OY	1883	ctgtctcttgagccgctcttgacacaagaaggtctcttgagccttaagcttgacacaggggaatgag	1942
Db	15672	CTTTCTTTGGCGTGTCGG---AAAAAAGGTTCCTGCTTACGCTGA--AAGGTGAAAGTGGG	15617
OY	1943	gcgcgcgaagagcgagcgaagcgcgaagagcgagcgctctccggatccagctccgccttgagc	2002
Db	15616	GGCGCCGGAAGGCGCGCGAAGCGGAGAGCGCGCGCTCTCCGGGATTCACCTCCGCCCTTGCG	15557
OY	2003	cagctgtgagcccgaggagctccctctgtgcccgcctgagagcgagagaacaagggagactga	2062
Db	15556	CAGTGTGGGGCCGGGGGGCTCCCTGTGTGCGCGGTGAGGCGAGAGAAACA--GGGACTTGA	15499
OY	2063	gtctcgaacagcagctgtgtcttcttaacttaattatcttagccctcagctccctgagctct	2122
Db	15498	GTCTCGAACACAGGGTGTGTTTTTACTTATTTATCTAGGGCCCTCAGCTCCTTGACGTCT	15439
OY	2123	gaagctccctcttgagcgtcttgagcgtctctctctgacactaagatgcagacaacagaagct	2182
Db	15438	GAGCTCTCCCTGTGAGCGCTGTGCGCTTCTCTCTGACACTGAGATGCGAGAACACAGACGCT	15379
OY	2183	tcggtgtgcttaaggagacagcaaaccttagaacccgcgcgctcttcggggaaaaacta	2242
Db	15378	TCGGGTGTGGCTTAGGGCAACAGCCAACTTAGAGACCAAGCGGGCTTTGGGGAAAA--A	15322
OY	2243	aagaagagacatctaaaatgttaatgtttaactagcttcaagataatctcttggaaaa	2302
Db	15321	AAGAAAGACATCTAAAAATGTTAAATGTTTAAAGCTTTCAAAATATTTATCTTGGAAAA	15262
OY	2303	atcaaggtcttgctgagacttgacactaatgtgtacagtaactctgtaactcttgacacacac	2362
Db	15261	ATCAGGGTTTGGCTGTGAGACTTGCACATATTTGTACAGTTAACTTGTGACTTTGACACACAC	15202
OY	2363	ctgaagatgacctccaacctcttgtaggctcttagagccttttaagaacctgtgtgagacctc	2422
Db	15201	CTGAAGATGCTCCACACTTGTGTAGGGCTTAGGGCTTTTATTCAGCCCTGGGTGGACCCC	15142
OY	2423	aggagccctctctctccctccctctctgtgtaactctctctgactctgtagaagatgctcta	2482
Db	15141	AGGGCCCTTCTCTTCCCTCTCCCTTCTGTGCTATTTCTCTGAGCTTGTAGAGAAATGTCCTA	15082
OY	2483	agaagatgtgactcaacagacgtctgagatccatgtgtccaatgaatgaatgagctatgagactg	2542
Db	15081	AGAAAGTGTGACTCACAGACCTGTGATTCATGTGTCAATTAATGAGCTGATGGAGACTGG	15022
OY	2543	agaaaggtctaataatccaatgagatcttgccctgtgttgccaattagaagccgagatgctc	2602
Db	15021	AGAAAGGCTTTAAATTCATAGGAGATC-TGCCGTGTGTGGCAATTTAAGGCGCGAGATGCTC	14963
OY	2603	gagagagat 2610	
Db	14962	GAGGAGT 14955	
RESULT 4	AP000517/c	AP000517 100000 bp DNA	PRI 30-MAR-2000
LOCUS	DEFINITION	Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region, section 16/20.	
ACCESSION	AP000517	AP000517	
VERSION	AP000517.1	GI:5926704	
KEYWORDS	SOURCE	Homo sapiens DNM.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.	
AUTHORS		Shima,S., Tamiya,G., Oka,A. and Inoko,H.	
TITLE		Homo sapiens 2,729,817bp genomic DNA of 6p21.3 HLA class I region	
JOURNAL		Published Only in Databank(1999) In press	

REFERENCE	FEATURES	COMMENT
2 (bases 1 to 100000) Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J. AUTHORS TITLE JOURNAL		
Submitted (21-Sep-1999) to the DDBJ/EMBL/GenBank databases. Miki Direct Submission Hirakawa, Japan Science and Technology Corporation (JST), Advanced Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081, Japan (E-mail:mikak@tokyo.jst.go.jp, URL:http://www-alls.tokyo.jst.go.jp/, Fax:81-3-5214-8470)		
		This sequence is conducted by Tokai University as a JST sequencing Team.
		Principal Investigator: Hidekoshi Inoko Ph.D Phone:+81-463-93-1121, Fax:+81-463-94-8884. The sequence is submitted by Human Genome Sequencing in ALIS project of JST
		Japan Science and Technology Corporation (JST) 5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan For further information about this sequences, please visit our sequence archive Web site (http://www-alls.tokyo.jst.go.jp/HGS/top.html) or send email to webmaster@www-alls.tokyo.jst.go.jp .
	Location/Qualifiers	
	1. 100000	
	/organism="Homo sapiens"	
	/db.xref="taxon:9606"	
	/chromosome="6"	
	/map="6p21.3"	
	complement(38199..38348)	
	/standard_name="D6S2198"	
	/note="SHGC-16144;The location is between each flanking site of PCR primers."	
	/db.xref="GDB:741200"	
	complement(45322..45412)	
	/note="906L;The location is between each flanking site of PCR primers."	
	/db.xref="GDB:1297693"	
	complement(52370..52547)	
	/standard_name="D6S2267"	
	/note="sy924b3-L;The location is between each flanking site of PCR primers."	
	/db.xref="GDB:15583975"	
	57335..57660	
	/standard_name="D6S1944"	
	/note="W1-7404;The location is between each flanking site of PCR primers."	
	/db.xref="GDB:679168"	
	57336..57491	
	/note="RH17987;The location is between each flanking site of PCR primers."	
	/db.xref="GDB:4572735"	
	complement(81490..81630)	
	/note="SHGC-17229;The location is between each flanking site of PCR primers."	
	/db.xref="GDB:1211045"	
	complement(88508..88638)	
	/note="SGC34813;The location is between each flanking site of PCR primers."	
	/db.xref="GDB:4504664"	
	/db.xref="GDB:4504664"	
	22059 c 22429 g 26811 t	
BASE COUNT	28701 a	
ORIGIN		
Query Match	37.2%; Score 971.2; DB 9; Length 100000;	
Best Local Similarity	94.4%; Pred. No. 3.2e-210;	
Matches 1140; Conservative	0; Mismatches 48; Indels 20; Gaps 12;	
OY 1403	aggaggtcattccttgagaccctcagccgagccgagccgagagcttgcttcgcggaagca 1462	
DB 70311	AGGGGCTCATCTACTGTGGACCTCTCAGACCGCCAGCGCGAG-CCTGGTTCTTCGGAGACA 70353	
OY 1463	ggaagtcagtaaggtacaccgcgcgagaagaagagccttcacagacagcccttcgccttcg 1522	
DB 70252	GGAGGTCAGTAGGTACACCGCGAGAGAGAGAGGCTTCCACAGACGCCCCCTTGCGCTTCG 70193	

QY	1523	aagagctcccgccggtgtctcgggtcccggtctctctccgggcccacagcttgcag	1582
Db	70192	ACGGCTCCCGGCGGCTTCTGGGCTTCCGGGCTTCTCTCCGGGCGCACCGCTGGCAGG	70133
QY	1583	cttaacctgacgtctggcgacgcccggcctbcaagcttggaggtgtgcggggaggggtga	1642
Db	70132	TTGACCTGCAGCTGGGGCGACGGCGGGCTCACGGTGGGGGTGCGCGGGAGGGGTGA	70073
QY	1643	ggaggaagggagagatbtagactcaagcgccgaggaacggcgtctggcgctgatcatctct	1702
Db	70072	GGAGGA - AGGGAGAGATGGAGACTCAGCCCGCAGAGACGGCTGTGGCGCGTATCTCG	70014
QY	1703	gaaacaagaatgctggggccagcacctcccgggcacacgacctgtccgtgagcgatc	1762
Db	70013	CACC - AGCAATGCTGTGGCCACACACTCCCGGGCCACCGAGCTG - CCGCTGACCGAGATC	69957
QY	1763	ccgcgcagacgctgagagctgcgacctbgaactacagagcgggaggtgaacctcaaacgc	1822
Db	69956	CGCGCG - GGGCGTGAAGTGGCTTCCTTGGACTACGAGCGGGGAGGTGACCTCCACACCG	69898
QY	1823	ccagagcccaagggcccacatctcaacctcaactagctctcttctccggccaaagttctcc	1882
Db	69897	CCAGA - CCCAGAGACCCATCTTACCTTACCTGACCTGCTCTT - CTCGGGCAAGCTTCC	69842
QY	1883	ctgtctcttggccgctctgacacaaaggctctgtgccttaagcttgacacggggaaatggg	1942
Db	69841	CTTTCTTTCCCTCTCG - - AAAAAAGTTCTCTCTCTTACCTGA - AAGCTTAACTGGG	69786
QY	1943	ggcgcggaagggccgacgaagcggaagcgagcctccgagatccagctcggccctgtgc	2002
Db	69785	GCCTCGGAAGGGCGGCGAAGCGAGACGGCGCTCTCGGATCCAGCTCGCCCTGGCG	69726
QY	2003	caagtgtcgagcccggggggctccctgtgcggcggtgagcgagagaaacagcggaacttga	2062
Db	69725	CAGTGTGCGGCGCGGGGGCTCCCTGTGCCCGGTGAGGCGAGAAACA - GGGGACTTGA	69668
QY	2063	gtctcgaaacagcgtgtgttttactttattatcttagccctcagctcctgaagctct	2122
Db	69667	GTCTCCAAACAGGGGTGTGTATTTACTTTATTTATCTTAGGCCCTCAGCTTCACTCTCT	69608
QY	2123	gaagccctcttggagcctctgacctctbgaactaagatgacagaacacgaagcgt	2182
Db	69607	GAGCTTCCTGTGAGCGCTTGCGCTTCTCTACCTTAGATGAGACACACAGAGGCT	69548
QY	2183	tggcgtgtgccttagggcaacagccaacctlaagaaacccgaccttgcgggaaaaacta	2242
Db	69547	TGCGGTGTCCTAGGGCAACAGCCAACTAGAGCCAGCGGGGCTTGGGGAAAA - A	69491
QY	2243	aagaagagacatctaaaaatgtaatgtttaactggttcaagaabaatacttgggaaaa	2302
Db	69490	AAACAAAACACTCTAAAAATAAATGTTAAACGTTTAAAAATTAATTACTTGGGAAAA	69431
QY	2303	atacaggttctbctgacttgacacttaattbtaagtaactcgttaacttggacaaac	2362
Db	69430	ATCAGGGTTTTCTGTGACTGTGACTTAATTTGTAAAGTTAACTTGTACACTTGTACACACAC	69371
QY	2363	ctgaagaatgcccaaccttggtagagcttlaagacctttatacagccctggtagagcccc	2422
Db	69370	CTGAAGATGCTGTCAACCTTTGTAGGGCTTAGGGCTTTTATCAGCCCTGGGTGACCCC	69311
QY	2423	aaggccctctccctccctccctctctctctctctctctctctctctctctctctctct	2482
Db	69310	AGGGCCCCCTTCCCTTCCCTCCCTCTGTGGTATTTCTGTGACTTGAAGATGTCTTA	69251
QY	2483	agaagaatgtagctacaagacctbgaatccaatgtgtcaaattagcgtatgtggacttg	2542
Db	69250	AGAAATGTGACTCAAGACCTGTGGATTCATGTGTCAATTAAGCCCTATGTGGACTGG	69191
QY	2543	agaaaggtctaaatcoaatbgaatctctgcctgttggcaattlaagccgaatgtgcc	2602
Db	69190	AGAAAGCTTAAATCAATGAGATC - TCCGTGTGTGGCAATTTAGGGCCGGAATGTGCTC	69132
QY	2603	gaaggagct 2610	

Db	69131	GAGGAGT	69124
RESULT	5		
LOCUS	AB023054/c		
DEFINITION	AB023054 Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region, clone:679G22, complete sequence.		
ACCESSION	AB023054		
VERSION	AB023054.1		
KEYWORDS	HTG.		
SOURCE	Homo sapiens DNA, clone:679G22.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (sites)		
TITLE	Shihna,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T., Kikana,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y., Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A., Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M., Bahram,S. and Inoko,H.		
JOURNAL	Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region		
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)		
REFERENCE	2 (bases 1 to 162383)		
AUTHORS	Shihna,T. and Takishima,N.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashima, Tokai University School of Medicine, Department of Molecular Life Science 2, Bohseidai, Isehara, Kanagawa 259-1193, Japan (E-mail:tsuhlina@tsl.ac.jp, tel:81-463-93-1121, fax:81-463-94-8884)		
FEATURES	location/Qualifiers		
source	1..162383		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="6"		
	/clone="679G22"		
	/map="6p21.3"		
	/label="6p21.3"		
BASE COUNT	45921 a 36561 c 36078 g 43823 t		
ORIGIN			
Query Match	37.28; Score 971.2; DB 9; Length 162383;		
Best Local Similarity	94.48; Pred. No. 2.9e-210;		
Matches 1140; Conservative	0; Mismatches 48; Indels 20; Gaps 12;		
QY 1403	aggagtcatactctggagccctcagaacccagacgagcagagacttggtttctcggagaaga 1462		
Db 63731	AGGGGTCTCTCACTCTGGACCTCAACACCGCCAGCCGGAG-CTTGCTCTCTGGAGAAC 63673		
QY 1463	ggaagtcagtaggtacaccccgagcaagaagagccttcacagacagcccccctgccttcg 1522		
Db 63672	GGAAGTCAGTAGGAGTACACCCGGGAGGAAGAGAGCTCCAGACAGGCCCCCTGCGCTTCG 63613		
QY 1523	acggatcctcggcgaggtcttgaggtctccggagcttcctcctcggcgccacagcgtgcag 1582		
Db 63612	ACGGCTCTCCCGCGGTTCTGGGCTTCCGGGCTTCTCTCCGGGCGCCACCCCTCGCAAG 63553		
QY 1583	ctgagctcagcctggcgacggcgagcgcgagcgtcagcgtgagggtgagcgagggaggtga 1642		
Db 63552	TTGACCTTCACCTGCGGCACCGCGGCGCTGCACGCTGGGGTGGCCGGGAGGGGGTGA 63433		
QY 1643	ggagagcagggagagatggagctcaagcgccgagagacggcgatctgggcccgtgatactct 1702		
Db 63492	GGAGGA-AGGAGAGATGGGACTCAGCCCGAGAGCGCGCTCTGGGCGGTGATCATCTCG 63434		
QY 1703	gcacacagcagtgctggtggcgacgacccctccggcgagccagcactgtgcgctagagagatc 1762		
Db 63433	CACC--ACCACTGTCTGGGCGCACACACTCCCGGGGCGACGACTG-CGCTGAGCGAGATC 63377		

QY 1763 ccgcgcagcgtgagatcgccttgactacagagcgagggtgacccccaacgc 1822
 |||||
 Db 63376 CCGGCG-CGGCTGAGAGTCGCCCTGAGACTAGAGCGGGGAGCTGAGCCCTCCCAACAGC 63318
 QY 1823 ccagagcccaaggcgccatcccttaaccttaactgctctttcccgcccaaggtctcc 1882
 |||||
 Db 63317 CCAGA--CCCAAGGAGCCCATCTTACCTTACCTGCTCTTT--CTCCGGCAAGTCTTCC 63262
 QY 1883 ctgtccttgccgccttgagacaaagggtcccttgaccttagctacagcgagggaatggg 1942
 |||||
 Db 63261 CTTTCTTTGGCGTCTGG--AAAAAGATTCTCCTTACCTGTA--AAGGCTGAAGTGGG 63206
 QY 1943 gcgcgcgaaggcgcgagaaagcgagacgagcgctcccgagatccagctccgcctggc 2002
 |||||
 Db 63205 GCGGCGCAAGGCGCGCAAGGAGAGCGGCGCTCCCGGATCCACCTCCGCCCTGGC 63146
 QY 2003 cagtgtcgagcccgagggtccctctgtcccgctgagcgagagaaacaacggaggaactga 2062
 |||||
 Db 63145 CAGTGTGCGGCCCGGGGGCTCCCTGTGCTCCGCTGAGGCGAGAGACAA--GGGACCTTGA 63088
 QY 2063 gtctgacagcggtgtgttttactatattatcttagagccctcaagctccctgaagctct 2122
 |||||
 Db 63087 GTCGGAACAGCGGTGTGTTTACTTATTTATCTAGGCCCTCAGCTCCCTGACGTCT 63028
 QY 2123 gagctccctgtgacgctcttgacctctctgacactagatgacagaaacagagcgct 2182
 |||||
 Db 63027 GAGGCTCCCTGTGACGCTGTGCTGCTCTGACCTGAGATGAGAACACACAGCGCT 62968
 QY 2183 tcgagctgtgcttaaggcaacagccactagaaaccccgagcctcttcgagggaacta 2242
 |||||
 Db 62967 TCGGCTGTGCTTAGGCGCAACAGCCATGAGACCGAGCGGCTTTGGGGAAAA--A 62911
 QY 2243 aagaagagacatctaaatgttaattgtaaaactgtttcaagataatattcttgggaaa 2302
 |||||
 Db 62910 AAGAAAGACATCTAAATGTTAACTGTTCAAAATATATCTTGGGAAAA 62851
 QY 2303 atcagagtttggcgagcttgactaattgttagagtttaacttgcttcttgaacacac 2382
 |||||
 Db 62850 ATCAGGGTGTGCTGAGCTGCACTAATTTGTACAGTTACTTGCTGTGAGACACAC 62791
 QY 2363 ctgaagatgctcccaactctgttagagcgcttaagcctttatcaagcctggtgagcccc 2422
 |||||
 Db 62790 CTGAGAGTGGCTCCACCTTTGTAGGCGCTTAGGCGCTTTTATACGCCCTGGGTGAGCCCC 62731
 QY 2423 agggcccttcccttcccttccctctctgtcaatcttctctgagactgttagaagatgtccta 2482
 |||||
 Db 62730 AGGGCCCTTCTCTTCCCTTCCCTTCTGTCATCTTCTGTGACTTGTAGAGAAATGTCCTA 62671
 QY 2483 aagaaggttactcagacactctgatactgctgctcaatgtgtcaattgagcgtgagtgactg 2542
 |||||
 Db 62670 AGAAAGTGTACTACAGACCTGTGATTCATGTGTCCATTTAGGCGCTGATGGACTGG 62611
 QY 2543 aagaagcttaaatcaatgagatgactgtgctgtgttgcaattagagcgcaagtgctc 2602
 |||||
 Db 62610 AGAAAGCTTAATTCATATGAGATC-TGCCGTGTGTGGCAATTAGGCGCGAGATGGCTC 62552
 QY 2603 gaaggaggt 2610
 |||||
 Db 62551 GAGGAGGT 62544
 |||||
 RESULT 6
 AB023055/c 162383 bp DNA PRI 20-NOV-1999
 LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
 DEFINITION clone:904J21, complete sequence.
 ACCESSION AB023055
 VERSION AB023055.1 GI:5672624
 KEYWORDS HTG.
 SOURCE Homo sapiens DNA, clone:904J21.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (sites)
 AUTHORS Shihina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
 Kikkawa,E., Iwata,K., Tomizawa,M., Okasaki,N., Kikano,Y.,
 Watanabe,K., Fukuzumi,Y., Itakura,S., Suwawara,C., Ono,A.,
 Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Sceda,E., Kimura,M.,
 Bahram,S. and Inoko,H.
 TITLE Molecular dynamics of MHC genes unraveled by sequence analysis of
 the 1,796,938-bp HLA class I region
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
 MEDLINE 20027539
 REFERENCE 2 (bases 1 to 162383)
 AUTHORS Shihina,T. and Takishima,N.
 TITLE Direct Submission
 JOURNAL Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi
 Shihina, Tokai University School of Medicine, Department of
 Molecular Life Science 2, Bohseidai, Isehara, Kanagawa 259-1193,
 Japan (E-mail: tshihina@is.tokai.ac.jp, Tel: 81-463-93-1121,
 Fax: 81-463-94-8884)
 FEATURES
 Location/Qualifiers
 1..162383
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="904J21"
 /map="6p21.3"
 BASE COUNT 45921 a 36561 c 36078 g 43823 t
 ORIGIN
 Query Match 37.2% Score 971.2; DB 9; Length 162383;
 Best Local Similarity 94.4% Pred. No. 2,9e+210;
 Matches 1140; Conservative 0; Mismatches 48; Indels 20; Gaps 12;
 QY 1403 aggggtcatcactctgagacctcagaccgacgagcgagacccgtgtctctcgaaagaca 1462
 |||||
 Db 63731 AGGGGTCTACTCTGTGACCTGAGACCGGACGCGAG--CTGGTCTCTCGGAAGACA 63673
 QY 1463 ggaagtcagtgtggttacaccgagaaagagagctgtcagacagccctgctctg 1522
 |||||
 Db 63672 GGAAGTCAGTGTGATACCCCGGAGAAAGCCCTGCCAGACGCCCTCGCTTCG 63613
 QY 1523 acggccctccggcggtgtctgtggtctccgggtctctcctccggcgagcagctgtgagag 1582
 |||||
 Db 63612 ACGGCTCCCGGCGGCTTCTGGCTCCCGGCTTCTCTCCGGGCGCACCGCTGGCAGG 63553
 QY 1583 ttgacctgacgttggcgagcgagcggtcgcaggtgtgggtgtggtcggagagaggtga 1642
 |||||
 Db 63552 TTGACCTGTGAGCTGGGCGAGCGGCGGCTGCAGGTGGGGGTGGCGGGAGGGGTGA 63493
 QY 1643 ggaagacagagagatgtgactcagcgccgagagagcggtctgtggcgtgatcatctc 1702
 |||||
 Db 63492 GGAAGA-AGGAGAGAGTGGGAGCTAGCGCGGAGAGCGGCTGTGGGCGGTGATCATCTCG 63434
 QY 1703 gacccaagagttctgtggcgagacactcccccgggacacagacactgtcgtgagcagatc 1762
 |||||
 Db 63433 CACC--AGCAGTGTGGGCCAGCACCTCCCGGACACGACCTG--CCGCTGACGAGATC 63377
 QY 1763 ccgcgcagcgtgagatcgccttgactacagagcgagggtgacccccaacgc 1822
 |||||
 Db 63376 CCGGCG-CGGCTGAGAGTCGCCCTGAGACTAGAGCGGGGAGCTGAGCCCTCCCAACAGC 63318
 QY 1823 ccagagcccaaggcgccatcccttaaccttaactgctcttttcccgccaaggtctcc 1882
 |||||
 Db 63317 CCAGA--CCCAAGGAGCCCATCTTACCTTACCTGCTCTTT--CTCCGGCAAGTCTTCC 63262
 QY 1883 ctgtccttgccgccttgagacaaagggtcccttgaccttagctacagcgagggaatggg 1942
 |||||
 Db 63261 CTTTCTTTGGCGTCTGG--AAAAAGATTCTCCTTACCTGTA--AAGGCTGAAGTGGG 63206
 QY 1943 gcgcgcgaaggcgcgagaaagcgagacgagcgctcccgagatccagctccgcctggc 2002
 |||||
 Db 63205 GCGGCGCAAGGCGCGCAAGGAGAGCGGCGCTCCCGGATCCACCTCCGCCCTGGC 63146


```

repeat_unit      /number=3
                  complement(43600..43880)
                  /gene="r1b30"
                  /rpt_family="sine"
exon              complement(44457..44687)
                  /gene="r1b30"
                  /number=3
                  complement(44688..45258)
                  /gene="r1b30"
                  /number=2
                  complement(45259..45354)
                  /gene="r1b30"
                  /number=2
                  complement(45355..48078)
                  /gene="r1b30"
                  /number=1
                  /number=1
                  /rpt_family="sine"
                  /note="(CAA)-7AA-(CAA)9"
                  /rpt_family="microsatellite"
                  complement(46810..47073)
                  /gene="r1b30"
                  /rpt_family="sine"
                  complement(48079..48510)
                  /gene="r1b30"
                  /number=1
                  /note="(CA)891
                  /number=1
                  /rpt_family="microsatellite"
                  join(51540..51908,53456..53551,57555..57785,58519..58541,
                  60253..60366,60696..60728,61939..62456)
                  /gene="znfb7"
                  /codon_start=1
                  /product="putative zinc finger protein"
                  /db_xref="GI:6625537"
                  /db_xref="SPTREMBL:O9TSMO"
                  /translation="MPLTPSHKGVASDCQRLDVAVTAAAGHTFGLPLPOMGA
                  QPSRYVLCVCEKEQTEPLVPLVPLGEPGEGYCEHGEKIFYPCNDAEFLVCR
                  EGPSHVAHVGFLDEAIOPIYRDLRGLEALITLERDIEEMKSGREOKLOVLAQES
                  KRKHVATPEFRLQOEIOEORILLARITELERQIWKRPDKYIKSLSEVATLCTOYKE
                  LKRCOOPASELLQDIVRVNNSRCEKTFTVSPPEAISPDLVKIKIDLRKILITPEMLRA
                  FSENLVHLETDGIVTLDPILASPSPVLSEDKRSVRYTRKQNLDPSPRFGPLPV"

Query Match      12.7% Score 332.4; DB 4; Length 152211;
Best Local Similarity 72.0% Pred. No. 2.5e-65;
Matches 605; Conservative 0; Mismatches 206; Indels 29; Gaps 12;

QY 1402 cagggtcatcactctcgtcagccctcagaccgagcgagagagctgtctctcggaagac 1461
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61936 CAGGCATCGTCACTCTGACCCCTCTGACCCGACCCGAG-CCGTGTCCTTCCGAGAGAC 61994
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1462 aggaagtcagtgagtgacacccagagaagaagagcctgcagagcccccctgcgcttc 1521
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61995 AGGAAGTCGTCGCGTACACTCGGCAGACAGAACTGCCCGACAGCCCACTGCGCTTC 62054
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1522 gacgacctcccgcggtctcgtgagcttcggggtctctcccgagcgagcccgctgcag 1581
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62055 GAGGCTCTCCGCGTGCCTGCGGCTCCCGGCTTCTCCTGTGGCGCCACCGCTGGCAG 62114
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1582 gtcgacctcagctggcgagcgagcgcgctgcaggtgaggtgaggtgaggtgaggtg 1641
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62115 GTGGAGTGCAGTCGAGAGAGGCGCGGCTGCACTGTGGGGTGTGCGGAGAGAGGTG 62174
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1642 aggaagagcagagagagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1701
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62175 AGGAGGA-AGGAGGAGCAGGCTGAGCGCGAGAGGCGGTGCGCGGTGAGTC--CTC 62231
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1702 tgcacacagagtgctgagcgagcactcccgagcgagcagcagctgtcagtgagagat 1761
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62232 TCCACACGAGTGTCTGGCGCAGCACTCCCGGCGACCGACCTG-CCGCTGAGCGAGAT 62290

```

```

QY 1762 ccgcgcagcgctgagagtgccctgagctacgagcgagcgagtgagctccacacag 1821
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62291 CCCGCGC-CGCGTGGGCGTGCCTCGACTACAGAGCGGCGCGCTGCGCTCTCTCAAGC 62349
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1822 ccgcgagcccgagggcccatcctcaccctcaccagtgctctcttcctccgccaagcttc 1881
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62350 CCGAGA--CCCGGGCGCCATCTTTCACCTTTCGCGGCTC-CTTTCGCGCAAAAGTGTTC 62406
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1882 cctgtccttgccgcctcgtgacacaaaggctcgtgagccttgaagcagggggaatg 1941
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62407 C-TTCTTCTGCTGTCTG--AAAAAGCTCCCTGCTGACTTGAAGCTGAGGGCA 62462
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1942 ggcgcgcgaaggcgagcgagagagagcgagcgctcctcgagatccagctccgcttg 2001
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62463 GCGTCGAGAGGCGCGTGCAGCGAGAGAGAGCAG-C-CGCGGATCCAACTCCACCTTG 62521
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2002 ccagtgctgagcccgaggggctccctgctgagcgagcgagtgaggaagaacagcgactg 2061
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62522 -----GGCGTTTCTTGTGCCCCAGACCGCAGGAGACGCGCTCAGAGT 62567
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2062 agtcgacagcgagtgcttcttacttatttatttctttaggcccctcagctccagctc 2121
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62568 CCGCAGTGTCTGTATTTTTCATTTATTTATCTGATGCCCT-TCAGCCTTGAGCTTC 62626
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2122 tgagctccctgtagagctctgctctctcagcctcagagtgagagacacagcgagc 2181
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62627 CTGGTTCCTCCGCTGACGCTCCGCTTCCTGCTGCACTCCGAGAGAGACGACGAGCAGA 62686
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2182 ttgcgtgctgctcagggcagcaacagcagcagcagcagcagcagcagcagcagcagc 2241
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62687 CGAGGCGCGCTGAGACAGCAGCAGCCTGGGAGCGCGGTGGGTTCGGGGAGAAATTT 62746
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
HS73B8R/c 318 bp DNA PRI 22-OCT-1995
LOCUS H.sapiens Cpg Island DNA genomic MseI fragment, clone 73b8, reverse
DEFINITION read cpg73b8.r1a.
ACCESSION 262790.1 GI:1035168
VERSION Cpg island; genomic MseI fragment.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 318)
AUTHORS Macdonald,M., Huckle,E., Wilkinson,P. and Micklem,G.
TITLE Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL CB10 IQO, England. E-mail contact: humquerry@sanger.ac.uk
REFERENCE 2 (bases 1 to 318)
AUTHORS Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
TITLE Purification of Cpg Islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
COMMENT Vector: pGEN-52f(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 IQO, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biophelp@hgmp.mrc.ac.uk.
FEATURES
source
1. 318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="male"
/dev_stage="adult"
/clone_type="blood"
/clone_lib="CGI-1"
/clone="73b8"
BASE COUNT 62 a 94 c 92 g 66 t 4 others
ORIGIN

```


BASE COUNT 42886 a 36022 c 35946 g 43560 t
ORIGIN

Query Match 7.6%; Score 198.8; DB 10; Length 158414;
Best Local Similarity 64.0%; Pred. No. 5.1e-35;

Matches 553; Conservative 0; Mismatches 217; Indels 94; Gaps 13;

```

QY 30 tatggtccaccatctgcaatctgacagagagcttcccccagttcccccagccac 89
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 35112 TAGGCTCTACCTGCGCCACGACGAGAGGGGCTCTCCCAATAGGCTCACCTTC 35053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 90 tgtgtctacagatcccaagaaa-cccaccctctgtacccaaggtg---tgtctgt 145
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 35052 TGTAGCCAAAGATTCACGAAACCCCATTTCCAGTACCTCATGTGCAATACCTGC 34993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 146 tctccaccctcagagagagagagagagagagagagagagagagagagagag 205
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34992 TCCCCACCCCAAGGACC-GAAGGCGACGAGGTGAGAGCTAGGTCCCTGCTGCTT 34934
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 206 ggaagaattctggttaattcagagagagagagagagagagagagagagagag 265
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34933 GGGAAAGCGGGGT-AGTCAAGAGCGCCCTGAGTTTGACCTGTTAGCTGGTTC 34875
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 266 gggcagaggttcttggaacaaagagcagagcagagtggtctccctgtgcatct 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34874 GGGC-----GTGGCAGCAGGTGACCTCGTGCAGCTGACATTTC 34836
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 326 tgtgtctctctctctctctctctctctctctctctctctctctctctctcagcc 385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34835 TGGCTTTCTTTTCAAGCTC-----CGATTCTTTGCTTGTGACAGTGAAGCA--- 34786
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 386 ttgcagagccgttcccccctctgctcagtgtagtgagtgagtgagtgagtgag 445
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34787 -----CTTAAAGTGAAGTCACTTTCAGGGAAGG 34759
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 446 gaactcgtctggtcgtcag-gagacggagtgagcggcgttggaagagacagtgatgcc 504
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34758 AAATTGAGAGCGACGCGTACAGAGACTGTGACAGCAGGGAAGCAGGTGATGCC 34699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 505 ggaaccccccctccctcgtgaaggggtgtcagtgagtgagtgagtgagtgagtgag 564
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34698 TCACCCGGTCC-----TGGGGGACTTTCACCAAGTACGCTGCTTCGACGCC 34647
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 565 ggcgcgtcgtgagatgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 624
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34646 GACCTTTCGAGATGGGTGACACAGACGCGGACACACTGTCTGCTTACGCTTC 34587
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 625 cgcgctctccagatgaggggcccacatcctgtgtgcaagatcctgtctgtgccgctctgc 684
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34586 ---CGGCGACCCCAATGGAGGCCAGCTGCTGTG-----CCCTCTCTGTCAAGGG 34540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 685 caagagagatgagcagcagagagagagagagagagagagagagagagagagagat 744
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34539 GCGAAGAGAGAGAAATCCAGGCCGCAAGTGCCCGGTGCTCCAGGTCCCTGAGCGA- 34479
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 745 aacttactcgaagagcagcagcagagagagagagagagagagagagagagagagat 804
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34480 GACCTTCTGAGAGAGAGAGAGAGAGATCTATTCTTC-TGTAAAGACAGAGCGGAGC 34422
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 805 tctctgtgtgtctgtcagagagaggttccacgacagcagcagcagcagcaggttctctgtg 864
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34421 TGTCTTGTGCGTGTCTGTGAGGAGGCGCCGCCACAGGACACTGTGGGTTCTCTGG 34362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 865 acgagagcattcagccctaccggg 888
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34361 ATGAAAGCATTCACACCTTACAGGG 34338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 12

AF220122 AF220122 2002 bp mRNA PRI 18-MAY-2001
LOCUS Homo sapiens tripartite motif protein TRIM10 alpha (TRIM10) mRNA.
DEFINITION

```

complete cds; alternatively spliced.
ACCESSION AF220122
VERSION AF220122.1 GI:12407412
KEYWORDS
SOURCE
ORGANISM human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2002)
AUTHORS Reymond, A., Meroni, G., Fantozzi, A., Merla, G., Cairo, S., Luzzi, L.,
            Riganelli, D., Zanaria, E., Messali, S., Calinca, S., Guffanti, A.,
            Minucci, S., Pellicci, P.G. and Ballabio, A.
TITLE The tripartite motif family identifies cell compartments
JOURNAL EMBO J. 20 (9), 2140-2151 (2001)
PUBMED 11331580
REFERENCE 2 (bases 1 to 2002)
AUTHORS Reymond, A. and Meroni, G.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) TIGEM, Via Olgettina 58, Milan, MI 20132,
            Italy
FEATURES
source location/Qualifiers
    1..2002
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    1..2002
    /gene="TRIM10"
    72..1520
    /note="TRIM10"
    /note="alternatively spliced"
    /codon_start=1
    /product="tripartite motif protein TRIM10 alpha"
    /protein_id="AA053495.1"
    /db_xref="GI:12407413"
    /translation="MASASVMSIADENVNCPICOGTLPREPVTIDGHNCRACLTAPYC
            EHPGPLESPFCPLCKEPPRGSPFRPMQIANVYENIRLDVSLTGLGEEVDYCEH
            GSKITFPCEDEDMOLCVCKREGEATHTMRLEDAAPVRYOLHKLCJLKEKEE
            IOELISRKRMQVLLTOVSTRKQOVIISFAHLRFLRQOOSILLAQLESODGDLRO
            RDEFLDVAIGEIRFESALIELEKERNERARELDIRSLTRCETRCRKPVAVSPE
            LQORIDPEQOALPQOREMKMFLKELFELDEPAHISLDPTSHPKLLSBDHQAQ
            FSSKRNPSDNDQREDFRATCVLAHLGTSFGTHRYVYSIDLAIGASCTGVSEVDQK
            GELRLRPREGVNAVRLANGFVSAISGFRPLRLKEDPRQVRSLODEVGMVFTNAVY
            REPITYFTASFRKVIYFPGMGSSSELS"
BASE COUNT 467 a 518 c 578 g 439 t
ORIGIN
Query Match 7.3%; Score 191.6; DB 9; Length 2002;
Best Local Similarity 55.9%; Pred. No. 4.4e-33;
Matches 651; Conservative 0; Mismatches 484; Indels 29; Gaps 14;
QY 739 ggaataacttactcgcgagagcagcagagagatcacttctcttcgcgaagcagt 798
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 356 GAGAGATGTGTGCCAAGAGCAGCAGAGAAAGATCTACTTCTTC-TGTGAGAGATGATG 414
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 799 ccgaagtcctctgtgtgtcttcgagagaggtcccaagcagcagcagcagcagcagcag 858
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 415 AGATGCAAGTTGTGCGTGTGTCGCCGGAGCGCTGGAGCACGCTACCAACCATGGCT 474
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 859 tcttgagagggccattcaagccctcagcgatgctcagagagtcgagagtcagagtcga 918
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 475 TCCCTGGAGATGACAGCGGCTCCCTATAGGGAGACAAATCATTAAGTGCTAAATGCTTAA 534
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 919 gcaagagagagatgagatgtlaaagtgtcaagaagacaggaagcttcaagt 978
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 535 GAAGAAGAGAGAGAGAGAT-CAAGAATTCAGTCAAGAAATAAAGATGCAAGTC 593
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 979 cgcgtcagcagatcgaacaagaagccgtcaggggtgacacagcctcttgagagc 1038
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 594 CTCCTACTACAGGTGTCCACACAGACACACAGTATTCTGA-----GTTCCGACAC 647
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1039 tgcgaagcggagctcagcagcagagatgctctctctgtcgcgagagatgagtgtaagc 1098
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 648 CTGAGGAAGTTTCTTAGAGGAACAGCAGAGCATCTCTTAGCACA-----ATTGAGAGC 701
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

QY	1099	tcggagtcacagatttggaagagagagatataatcacaaagctctcgaagatc	1158
Db	702	CAGATGGGACATCTTAGGCAACGGATGATTTGATTGGTGGTGGAGATC	761
QY	1159	aaccggttgagcccaagctcaagagagctcgagagagaatgtcagcagaagtg	1218
Db	762	TGCGGTTAGTCTTATTGAAAGAA--CTGGAGAGAGAAATGACAGCCAGCAAGG	819
QY	1219	agctctacaagatgtcagatcaagccagcggtgtgtgagatgaacttttgagt	1278
Db	820	AGCTCTGACGACATCAGAGCACTCTA-ATAAGATGTGAACAGAAAGTGGGAAA	878
QY	1279	cctgaagccattctccctgacccgtgtcaagaagatccgtatttcccaagaaatact	1338
Db	879	CCGGTGGCTGTGTGGCCA-GAGCTGGGGCAGAGATTTGGGACTTTCCCGACAGCCCT	937
QY	1339	caaccctccagagatgatgaatgtctcaagaaacttggcgatcatctggaatag	1398
Db	938	CCCCCTGACAGGAGATGAGATGTT-TCGTGAATAACTATAGCTTTGAGTTGACTATG	996
QY	1399	attcaggggtatcaactcttgagaccctcagaccgacccgagacccgtgtctctcgaa	1458
Db	997	AGCCAGCTCACATTCTCTAGACCTCAGACTTCCACCCCA-AGCTCCTTGTGCCAG	1055
QY	1459	gacaggaagtcatgtaggtacacccgagagaagagcctgcagacagccctgcgc	1518
Db	1056	GACACACGACGACCTCACTTCTCCCAAAATGGCAGAACTCACCGACACACCCACGCT	1115
QY	1519	tctgaagcctcccgcggtctcgtggtctccgggtctctcctccggcgccacccgtcg	1578
Db	1116	TTTGACGGGCGCACCTGCTTGTGGCCACACTGGCATCACAGGGGGAGACACAGTGG	1175
QY	1579	caggttgaccgctgctgagcgagcgcggtctgacaggtgtgggtgtgacggagaggg	1638
Db	1176	GTGGTAGTATAGACTCTGGCCATGGGCGAGCTGACCGTGGGGGTGAGCAGGAT	1235
QY	1639	gtcggaagagagagagatgagactcagcgccgagagcgagcgtctcgttatcat	1698
Db	1236	GTGACAGCGA-AGGGGAGCTTGGGCTGGGCGACAGAGAGGGGCTGTGAGGCT	1294
QY	1699	ctctgcacacaagcagtgctggtgcagacccctcccgagacagcagctgtccgtgagca	1758
Db	1295	GCTTGGGGCTTCTGGGCTTCCACACAGGGCTGACCTGTAAGAG	1352
QY	1759	gattcccgagcggtgtgagatgcctcgtgactacgagagcggtgaggttgaccctccaca	1818
Db	1353	CAGCCCGGCGAG-GTGAAGGTGTCTTTGACTATGAGGTGGGCTGGGTGACCTTACCA	1410
QY	1819	agccccaagagccaggggacatctcctacatcagctgtcttctcccgccaaggtc	1878
Db	1411	ACGGCTGACCCGAGAG--CCCATCTACACCTTCACTGCCCTTCACTAGS-AAAGTC	1466
QY	1879	tccctgtccttggtgcgcctgac	1902
Db	1467	ATTCCCTTCTTGGGCTGTGGGGC	1490
RESULT	13		
AF220121			
LOCUS	2269 bp	mRNA	18-MAY-2001
DEFINITION	Mus musculus tripartite motif protein TRIM10 (Trim10) mRNA,		
ACCESSION	AF220121		
VERSION	AF220121.1	GI:12407410	
KEYWORDS			
SOURCE	house mouse,		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 2269)		
AUTHORS	Raymond,A., Meroni,G., Fantozzi,A., Merla,G., Cairo,S., Luzi,L.,		
	Riganelli,D., Zanaria,E., Messali,S., Calnarca,S., Guifanti,A.,		

TITLE	Minucci,S., Pellicci,P.G. and Ballabio,A.		
JOURNAL	The tripartite motif family identifies cell compartments		
PUBMED	EMBO J. 20 (9), 2140-2151 (2001)		
REFERENCE	11331580		
AUTHORS	2 (bases 1 to 2269)		
TITLE	Raymond,A. and Meroni,G.		
JOURNAL	Direct Submission		
FEATURES	Submitted (30-DEC-1999) TIGEM, Via Olgettina 58, Milan, MI 20132, Italy		
source	Location/Qualifiers		
gene	1..2269		
CDS	/organism="Mus musculus" /db_xref="taxon:10090" 1..2269 /gene="Trim10" 293..1762 /gene="Trim10" /codon_start=1 /product="tripartite motif protein TRIM10" /protein_id="AAG53494.1" /db_xref="GI:12407411" /translation="MASAPSTSLADVNCPIQGLREPTIDCGNFCRGCLTRIC ETIPBESLUSCLCKEPRPESFRFMOLANVENIERJLQSLGVEDEDCPEH GEKTYFCEDEADLCVARETQGHGATVFEFLDAGAPYREJLOKLVLRKERBI QETQSENKRLOVLLTQVATKRQOVIQSFANLSQFLQOQOTLALQLEGGDLKQOL EEFSLATGEICRSTLIELEENKRTFAGLLDISTRCTRCRPEALSPEL GORLRPPPOOAIPLRQMKTFLEKCELDYEPANI SLDPOTSHPKLLSDHRRARF SYKQNSNPDPQORDRYTCVLAOCGFLPFGHRTVMVNDLHAGSCYGVREYDRRG ELRRPREGITAMVLAAGFVSALSFPTRLALEQRPKQVSLDYEVGNITFVAATVQ EHITTFASFQKTFLEPLFGMLGRSSFSLSQCBGAVSL"		
BASE COUNT	572 a 546 c 627 g 524 t		
ORIGIN			
Query Match	6.5% Score 170.4; DB 10; Length 2269;		
Best Local Similarity	54.7% Pred. No. 2.8e-28;		
Matches 638; Conservative 0; Mismatches 501; Indels 27; Gaps 14;			
QY	721	gtgcccctgggccccttgagagataactatctgcagagagcagcgagaaatctact	780
Db	556	GTCCAGCGAGGCTGTGAGGTGGAGATGCTGTCCGGAAACGGGGAGAAATCTACT	615
QY	781	ctctctgagagacagctccgagctcctctgtgtgtcttcgagggaggggtccacagcacc	840
Db	616	CTT-TTGGGAGAGAGATGAGGACAGATTGTGTGTGATGCGCGCAACCGACAGCATG	674
QY	841	aggcgacacacggtgggtgtcctcgagagagccattcagccctaccggagatcgtcaga	900
Db	675	GGGCTCACACTGTGCGCTTCCGTGGAGAGCGACGAGGTCCCTACAGGGAACAAATACGA	734
QY	901	gtcgacttgaaagcttgagcaagagagatgagatgtgagatgaagtgaagaa	960
Db	735	AGTGTCTGTGTCTTAAGAAAGAGAGAAATTA-CAAGAAACCCAGTCAAGAGAA	793
QY	961	gaccgaagcttcaagctcgagctgacatcagatcgaacaagcagaagccgttaaggtga	1020
Db	794	AACAAAGAAATACAGTGTCTGTGACTGAGTGG--CCACCAAGAGGCAACAGGTGATCT	851
QY	1021	cacagctcccttgagagagctgcaagcggtgcagcagcagcagatgtctcctgtggcg	1080
Db	852	CACAGTTTGGCATCTGAGCCA---GTTCCTGCGACGACACAGACAGTCCCTTACCA	907
QY	1081	cagagctgagtgatcgctcggagtcagaagatttggaagagagagatgaatataatca	1140
Db	908	CA-----ACTGGAGGGGTTAGACGGGGACATCTGAAGACACAGAGAGATTATTC	961
QY	1141	aagctctgaggaagtaaccggttgagcccaagctcaagagagctcgtgagagaaat	1200
Db	962	CTGGCCACTGGGGAGATCTGCCGTTTACAGACCCGATTGAGAGACT--GAGAGAGAGA	1019
QY	1201	gtcagcagccagcaagtgaagcttctacaagatgtcaagatcaagcagaagcaggtgtag	1260
Db	1020	ACAAGAGAGACAGCCAGGGGCTCTGACGAGATATCAGAAAGCACTCTA-ATAAGATGTGA	1078

QY 1261 atgaagacttttggagtcctgaagccattctccctgacctgttcaagaagatccgtga 1320
 Db 1079 ACCGAAAGTGCAGGAAACCCGAGGCTATATC-CCGAGACTGGGCAAAAGATCCGGGA 1137
 QY 1321 ttccacagaagaataactacccctcccaagatgatagaaatgttctcaagaacttgg 1380
 Db 1138 CTTCCCCCAACAGGCCATCCCTCTGGCGAGGAGATGAAGACATT-TCGTGAAAACCTCT 1196
 QY 1381 cgaalcactggaatgatattcaagggtgtcactcgtgaacctcagcccgagccgga 1440
 Db 1197 GTTTGAGTTGGACTATGAGCCAGCTCATATCTCTAGACCCCAAGACTTTCCACCCCA 1256
 QY 1441 gaacctgtcttcgagaagaagaagtcagtgagatacaaccggcagaagaagagctg 1500
 Db 1257 -ACCTCTCTCTCTCTGAAGACACACGAGGCTCGTTCTCTCAAAATGGACGAATTCA 1315
 QY 1501 ccagagacccccctgctcctcgaagcctcccgcgcttctggagctcccgagcttctcc 1580
 Db 1316 CCAGATATCTCCACAAACCTTTGATCGAGTCACTGTCTTGCGCCCAATGTGTTTACA 1375
 QY 1561 tccgagcgacacgcgctgagagttgacctgaagctgagcgagcgagcgctgcaggtg 1620
 Db 1376 GGGGGGAGACACACTTGGATGGTGAATGTGACTTGGCCCATGAGGACAGCTGCATCTA 1435
 QY 1621 ggggtgagcgagggaggggtgaagagagacagagagatgagactcagcgccgagagcg 1680
 Db 1436 GGTGTGTGTAGAGAGATGT-ACGACGGAAGGAGAAATTGAGCTGAGACACAGAGAGGG 1494
 QY 1681 cgtctggcgctgatactctctgacacaagagtgctggcgacactcccgagcgacc 1740
 Db 1495 AATATGGGAGTGAGGCTGTGGGCTTGTGTTGACGCTCTGGGTTCTTCCCAACAG 1554
 QY 1741 gacctgctcgtgagcagagatcccgcgagcgagcgagagatgagcctgagtagagcg 1800
 Db 1555 G-CTGGCCCTGAGAGAGCAACCCGGGAG--GTGCAAGTATCTCTTACTATAGGTAG 1610
 QY 1801 ggcaggtgaccttccacaagcgccagagcccaaggggccacttcaacttactgctc 1860
 Db 1611 GCTGGATACACTTCTGTCATGCTGTCACTCAGAG--CATATCATATCACTTCACTGCTC 1668
 QY 1861 ttcttcgcgcaaggtcttccctgt 1886
 Db 1669 CTTTACACAGAAATCTTTCCTTGT 1694

RESULT 14
 AF134811
 LOCUS AF134811 2214 bp mRNA ROD 03-MAY-1999
 DEFINITION Mus musculus hematopoietic RING finger 1 (Herf1) mRNA, complete cds.
 ACCESSION AF134811
 VERSION AF134811.1 GI:4731627
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 2214)
 Harada,H., Harada,Y., O'Brien,D.P., Rice,D.S., Naewe,C.W. and Downing,J.R.
 TITLE A novel hematopoiesis-specific RING finger protein, is required for terminal differentiation of erythroid cells
 JOURNAL Mol. Cell. Biol. 19 (5), 3808-3815 (1999)
 MEDLINE 99232609
 REFERENCES 2 (bases 1 to 2214)
 Harada,H., Harada,Y., O'Brien,D.P., Rice,D.S., Naewe,C.W. and Downing,J.R.
 TITLE Direct Submission
 JOURNAL Submitted (14-MAR-1999) Pathology and Laboratory Medicine, St. Jude Children's Research Hospital, 332 North Lauderdale, Memphis, TN 38105, USA
 FEATURES
 Location/Qualifiers

source 1. .2214
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /tissue_type="spleen"
 gene 1. .2214
 /gene="Herf1"
 288. .1757
 /gene="Herf1"
 /note="Herf1"
 /codon_start=1
 /product="hematopoietic RING finger 1"
 /protein_id="A4D28534.1"
 /db_xref="GI:4731628"
 /translation="MASAPSVSLADEVNCPIQGTLRBPVYIDCGHNCRCGLTRVC
 EIRPESSESLCPLKEPRPSPSPNNOLAVNENIRLQASTRGLEVEDAPCEH
 GERTYPECEDEAOICVVCREMGHCAHVHFLFEDAAGYREIOKCYLCRKREKEI
 OENOSRENKRIQVILTOVAPKROOVISOFAHLSQFLOQOFTALQLEIDGDIKKOQ
 EERDSIATSEICQFSTLIELEBKMRTRAGLTDIRSLIRETKCKRPAISBEL
 GQRIRDPPOQALPDKQEMKTFLEKLCFLDYDEPAHISLPQTSHPRLILSEDIRRAF
 SYKMONSPDTPORFVTCVLAOCFTGRTWVNVVAVDAGSGSTGVAVREDBRRKG
 ELRLRDEEIMAVRLAMFVSALGSEPTRLALEOPRKYQVSLDYEVGWTVEVNAVTO
 EHIYTTASPTOKIIFPLFGIMGSGSESLSCQGAVALS"
 BASE COUNT 538 a 538 c 618 g 520 t
 ORIGIN

Query Match 6.3%; Score 165.6; DB 10; Length 2214;
 Best Local Similarity 54.5%; Pred. No. 3.4e-27;
 Matches 635; Conservative 0; Mismatches 504; Indels 27; Gaps 14;

QY 721 gtgcccttgagccgctgtagagataacttctgagagagacggcgagaagatcactt 780
 Db 551 GTCACGCGAGGCTGTGAGGTGAGATGCTGTGCCGAACAGCGGAGAAATCTACTT 610
 QY 781 ctctcttgagagacgagatgccgagttctctgtgtcttcagagagaggtcccaagacc 840
 Db 611 CTT-TTGGAGAGAGATGAGACACAGTTGTGTGTATCGCGGAACCGGACAGCATG 669
 QY 841 aggcgcacacgltgggttctctgagagacatcgaagccctacccgagatcgtctcaga 900
 Db 670 GGGCTACACTGTGCACTCTCTGAGAGACAGCAGAGTCCCTACAGGAAACAATAACAGA 729
 QY 901 gtgcagctggaagctctgagacagagagagatgagattgtgagatgtcaagaa 960
 Db 730 AGTGTCTCGTGTCTTAAAGAAAGAGAGAGAGAT-CAAGAAACCCAGTCAAGAGAA 788
 QY 961 gaccagaagcttcaagtgcgcgctgaactcagatcgaaacagcagagcgatgtctctgtgca 1020
 Db 789 AACAAAGAAATACAGTGCTCTGACTCAGGTGG--CCACCAAGAGCAACAGGTGATCT 846
 QY 1021 cacagctccttgagaggtctgaagcgagagctgacgacgacgagatgtctctgtgagc 1080
 Db 847 CACAGTTGGCGATCTGAGCA----GTTCCTGAGCAACAGCAGACTGCCCTTAAAGCA 902
 QY 1081 caggaactgagtgtaagctcgtgagatcacagatttgaagagagagatgaatatcaca 1140
 Db 903 CA-----ACTGGAGGGGTGAGACGGGACATCTTGAAGCAGCAGGAAAGTTGATTC 956
 QY 1141 aaggtctctgagaaagtcaccggcgttggagccccaagctcaagagagctggagagagagt 1200
 Db 957 CTGGCCACTGGGGAGATCTCCAGTTGACACCCCTGATTAGAGAGC--TGGAGAGAGAGA 1014
 QY 1201 gtacgagcagcagaagtgagctctcaaaagatgtaagagtcgaagccagagcaggtgag 1260
 Db 1015 ACAAGAGGACAGCCAGGGGCTCTCTAGCGATATTCAGAACACACTCTA-ATAAGATGTGAA 1073
 QY 1261 atgaagactttgtgagtcctgagagccattctccctgacctgttcaagaagatccgtga 1320
 Db 1074 ACCGAAAGTGCAGGAAACCCGAGGCTATATC-CCGAGACTGGGCAAAAGATCCGGGA 1132
 QY 1321 ttccacagaagaataactacccctcccaagatgatagagaatgttctcaagaacttgg 1380
 Db 1133 CTTCCCCCAACAGGCCATCCCTCTGGCGAGGAGATGAAGACATT-TCGTGAAAACCTCT 1191

```

QY 1381 cgcacatctgaaatagattcagaggtcactctgaccctcagaccgacgga 1440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1192 GTTTGAGTTGACTATGAGCAGCTCATATCTCTAGACCCAGACTTCCGCCCA 1251
QY 1441 gacctggtctctcgaagacaggaagtcagtgatcacccgacgaagaagcctg 1500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1252 -AGCTCCCTCTCTGAGAGCACCAGGGCTGCTCTCAATAATGACAGATTCA 1310
QY 1501 ccagacagcccccctgcgtcttcagagcgtcccgcggtcttcgggtctcc 1560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1311 CCAGATCTCCACAAACGCTTGTATGAGTCACTGTGTGCGCCAAATGTTTACA 1370
QY 1561 tccgagcgccacgcgctgcaagcttgacgtcagctgcgagcgcgctgcagctg 1620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1371 GGGGGGAGATACACTGGGTGCTGATGTATGACTTGGCCCATGAGGCACTGCTTA 1430
QY 1621 ggggtgagcgaggaaggggtgaggaagacagagagatggagactcagcgccgaagacg 1680
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1431 GGTGTGTCAGAGAAGATGT-ACGACGAGAGGAGAAATTGAGGCTGAGACACAGAGAGG 1489
QY 1681 cgtctggcgctgatactctgacccaagcagtgctggcgacactcccgagc 1740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1490 AATATGGGCAATGAGGCTGGCTGGGCTTGTTCACCTGCTGCTTCCACACAG 1549
QY 1741 gacctgtccgtgagcgagatcccgcgagcgctgagagctgcgtgactacagagcg 1800
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1550 G-CTGGCCCTGGAGGAGCAACCCCGAGG--GTCAGAGTATCTTACATGTAGGATG 1605
QY 1801 ggcaggtgacctccacaagcccaagagcgccactcactcactcactgctc 1860
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1606 GCTGATTAACCTTGTCAATCTGTCACTCAGAGG--CATATCTATACCTTCACTGCTC 1663
QY 1861 ttctccgagcgaaggtcttcctgt 1886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1664 CTTTACACAGAGATCTTCCCTGT 1689

RESULT 15
LOCUS   RM499P20      54674 bp      DNA      HTG      10-AUG-2001
DEFINITION   Rattus norvegicus clone RPCI-31-499P20 strain Brown Norway, ***
SEQUENCING   IN PROGRESS ***, in unordered pieces.
ACCESSION   AL603720
VERSION     AL603720.1 GI:15149582
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 54674)
            Sudbrak,R., Borzym,K., Mueller,I., Klages,S., Koslura,A.,
            Walter,L., Guenther,E., Hurt,P., Lehnach,H., Himmelbauer,H. and
            Reinhardt,R.
            Unpublished
JOURNAL     2 (bases 1 to 54674)
AUTHORS     MOGENR.
REFERENCE   Direct Submission
            Submitted (10-AUG-2001) MPIMG, Abt.Lehnach, Max Planck Institut
            fuer Molekulare Genetik, Inestrasse 73, Berlin, 14195 Germany
COMMENT     contig 01      1. 789
            contig 02      890. 1444
            contig 03      1545. 2322
            contig 04      2423. 3319
            contig 05      3420. 4782
            contig 06      4883. 7087
            contig 07      7188. 9565
            contig 08      9666. 12686
            contig 09      12787. 20768
            contig 10      20869. 25718
            contig 11      25819. 27038
            contig 12      27139. 27737

```

```

contig 13      27838. 35931
contig 14      36032. 37073
contig 15      37174. 37354
contig 16      37455. 37712
contig 17      37813. 38208
contig 18      38309. 38532
contig 19      38633. 38773
contig 20      38874. 38973
contig 21      39074. 39244
contig 22      39345. 39784
contig 23      39885. 39971
contig 24      40072. 40645
contig 25      40746. 45201
contig 26      45302. 45854
contig 27      45955. 47332
contig 28      47433. 48465
contig 29      48566. 49192
contig 30      49293. 50070
contig 31      50171. 50830
contig 32      50931. 51433
contig 33      51534. 52404
contig 34      52505. 52955
contig 35      53056. 53762
contig 36      53863. 54674.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES             Location/Qualifiers
     source           1..54674
                     /organism="Rattus norvegicus"
                     /strain="Brown Norway"
                     /db_xref="taxon:10116"
                     /clone="RPCI-31-499P20"
                     /clone_11b="RPCI-31"
                     /note="RT region on chromosome 20"

BASE COUNT   13633 a 12253 c 12039 g 13225 t 3524 others
ORIGIN
Query Match      6.2%; Score 160.6; DB 2; Length 54674;
Best Local Similarity 69.3%; Pred. No. 2.8e-26;
Matches 328; Conservative 0; Mismatches 119; Indels 26; Gaps 7;

QY 419  tgaagtactgatttaaggaaggaagacgcgtgtgctgag---gagacggagtg 475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48280 TAAAGTACTAGTTTCAAGGAAAGAAATGTGTGAGAGACAGACAGATGCGTG 48221
QY 476  gacggctgggaagacacgcgtgacgcccgaaccccgctccctggaagggtgtcca 535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48220 GACAGGACAGAGAAAGCAGCGGATGATCCCTTAACCGGATCCCTGGC-----GGGACC 48169
QY 536  tgaagtccctgctgacctctgtcgcgggcgcttgagagatgctggtacattccctg 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48168 CCACCAAGTAGCCTAGTGTGACACCGGCCCTCGACAGATGCGGTACATAGCTCG 48109
QY 596  tggacacacctctgcggctctgctcccgcgctctcccgataggggggcccatctc 655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48108 TGGACACGCTGTCTCTACTCTGCTCC---CGCGCCCAATGAGGCGCCACCTGCT 48052
QY 656  gtggcaagatccctgctgcgcgctctgcgaagagagtagcagagcagagactccatgg 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48051 GTC-----CCGCTCTGTGAG---GGGGCAGAAAGAGAGAAATCCAGACCGCAGTCA 48002
QY 716  cccctgtcccttggcgccgctggagataactactgcgagagacagcgcgagaagatc 775
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48001 CTCGGTATCCCTAGAGTCCCTGACGCA-GACCTGCTCGAAGACAGCAGAGAAATC 47943
QY 776  tactcttcttgaggaagcagatgcgagttctctgtgtgtcttcaggggggtccac 835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47942 TATTTCTTC-TGCAAGACAGACGCGAGCTGCTGTGCTCTGACGAGAGGCGCCGC 47884
QY 836  gaccacgcacacgcgtgggtctcttgacgagggcattcagccctaccgg 888

```

|||||
Db 47883 CCACCAGCACACACTGTGGGTTCTGTGATGAAGCCATCCACCTACAGGG 47831

Search completed: March 30, 2002, 13:47:16
Job time: 24318 sec